(12) UK Patent Application (19) GB (11) 2 302 199 (13) A

(43) Date of A Publication 08.01.1997

(21) Application No 9619932.8

(22) Date of Filing 24.09.1996

(71) Applicant(s)
AllVoice Computing plc

(Incorporated in the United Kingdom)

Courtenay Park, NEWTON ABBOT, Devon, TQ12 2HB, United Kingdom

(72) Inventor(s)
John Christopher Mitchell
Alan James Heard
Steve Norman Corbett
Nicholas John Daniel

(74) Agent and/or Address for Service
Beresford & Co
2-5 Warwick Court, High Holborn, LONDON, WC1R 5DJ, United Kingdom

(51) INT CL^6
G06F 17/21 , G10L 3/00

(52) UK CL (Edition O)
G4R RHA RHB R1F
G4H HTAD H1A H13D H14A H14D
U1S S2126

(56) Documents Cited
GB 2230370 A GB 2088106 A EP 0649144 A1
EP 0077194 A1 WO 93/07562 A1 US 5031113 A

(58) Field of Search
UK CL (Edition O) G4H HTAD HTAT , G4R REX RHA RHB
INT CL^6 G06F 3/00 3/16 17/00 17/20 17/21 17/22 17/24 , G10L 3/00 5/00 5/06 7/00 7/08 9/00 9/08 9/18
Online:WPI

(54) **Text processing**

(57) Data processing apparatus receives recognition data from a speech recognition engine and its corresponding dictated audio data. A display displays the recognised words or characters and the recognised words or characters are stored as a file together with the corresponding audio data. Link data is formed to link the position of the words or characters in the file and the position of the corresponding audio component in the audio data. The recognised words or characters can be processed without loosing the audio data. An audio message may be stored associated with a document.

The apparatus may display to the operator recognised words or characters which have a likelihood indicator below a preset threshold.

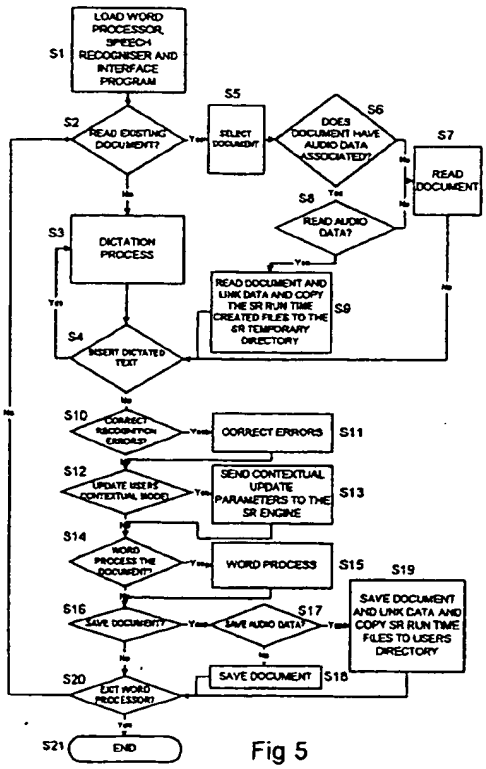


Fig 5

GB 2 302 199 A

At least one drawing originally filed was informal and the print reproduced here is taken from a later filed formal copy.

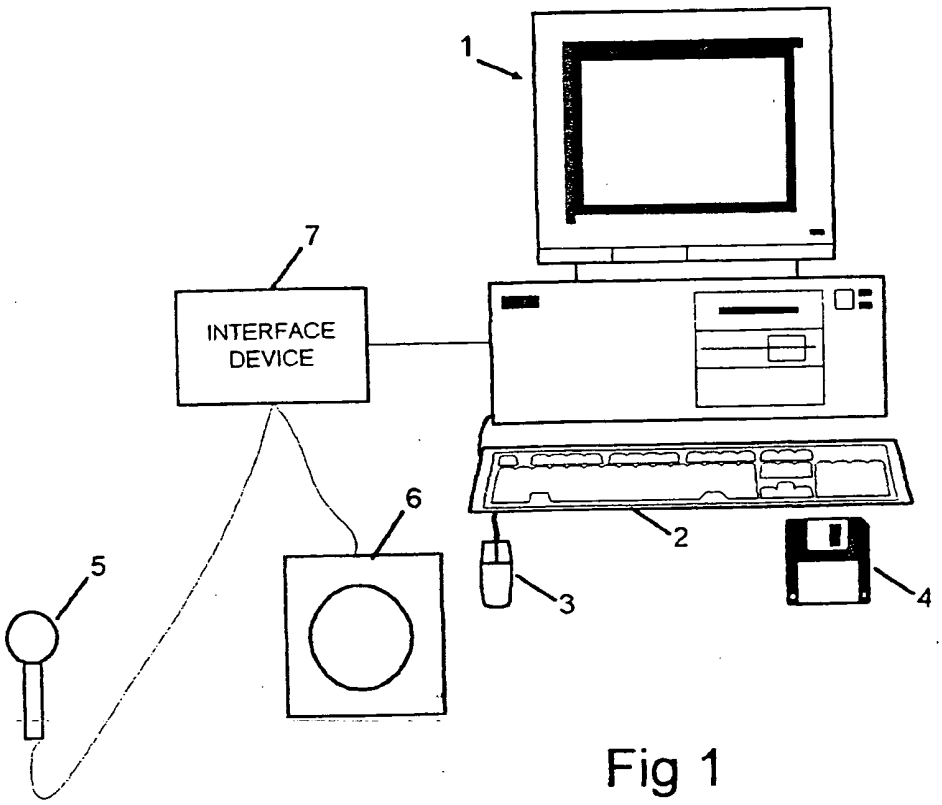


Fig 1

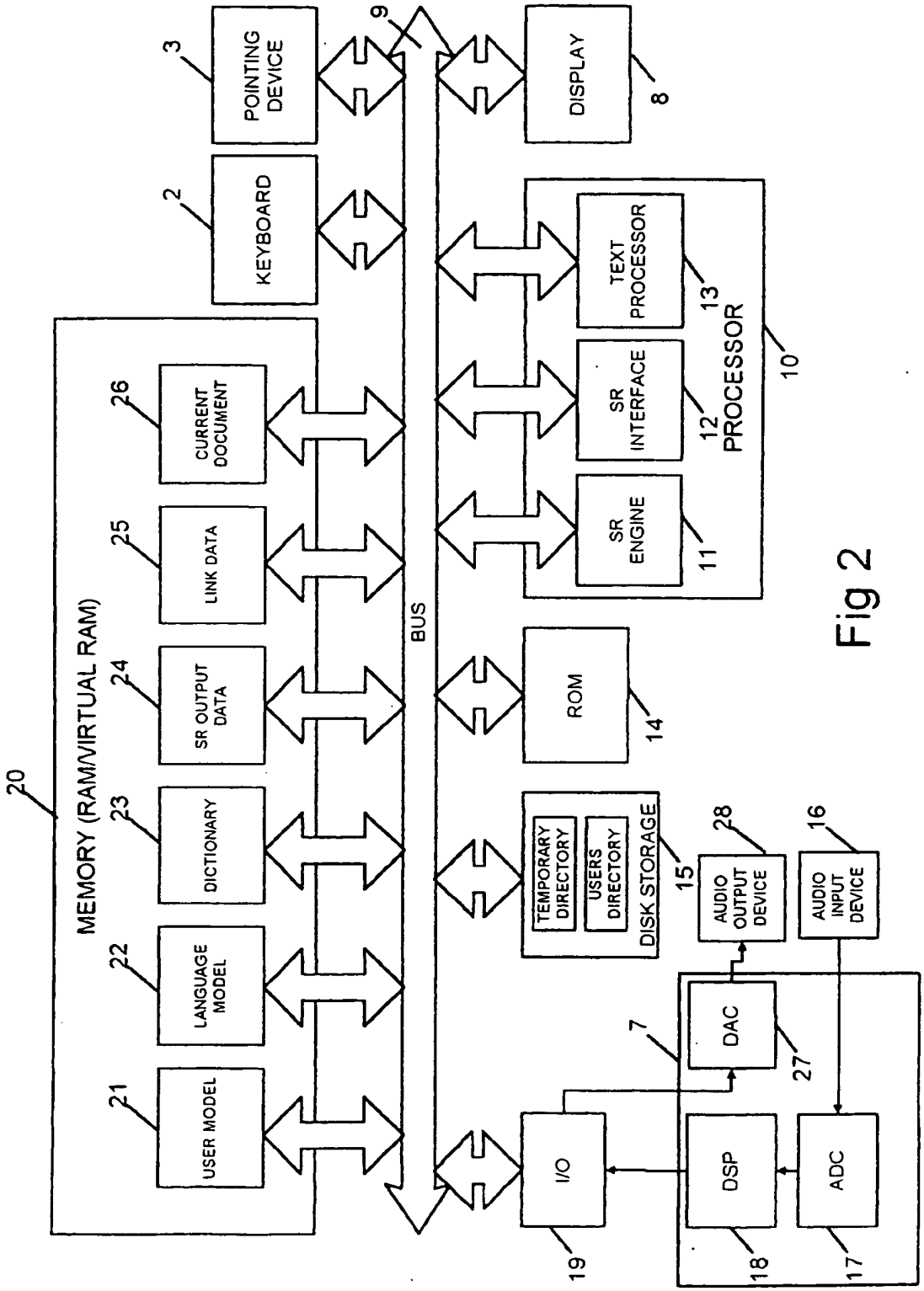

POINTING DEVICE
3
KEYBOARD
2
9
BUS
DISPLAY
8
MEMORY (RAM/VIRTUAL RAM)
20
CURRENT DOCUMENT
26
LINK DATA
25
SR OUTPUT DATA
24
DICTIONARY
23
LANGUAGE MODEL
22
USER MODEL
21
TEXT PROCESSOR
13
SR INTERFACE
12
SR ENGINE
11
PROCESSOR
10
ROM
14
TEMPORARY DIRECTORY
USERS DIRECTORY
DISK STORAGE
15
AUDIO OUTPUT DEVICE
28
AUDIO INPUT DEVICE
16
DAC
27
7
I/O
19
DSP
18
ADC
17

Fig 2

Identifier Tag	Audio Start	Audio End	Alternative words and Scores
1	0	199	$\Pi_{1a}W_{1a}$ $\Pi_{1b}W_{1b}$ $\Pi_{1c}W_{1c}$ .....
2	199	324	$\Pi_{2a}W_{2a}$ $\Pi_{2b}W_{2b}$ $\Pi_{2c}W_{2c}$ .....
3	324	361	...
4	361	...	
5	...		
...			

Fig 3

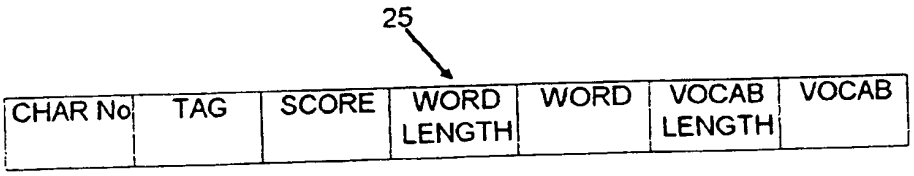


Fig 4

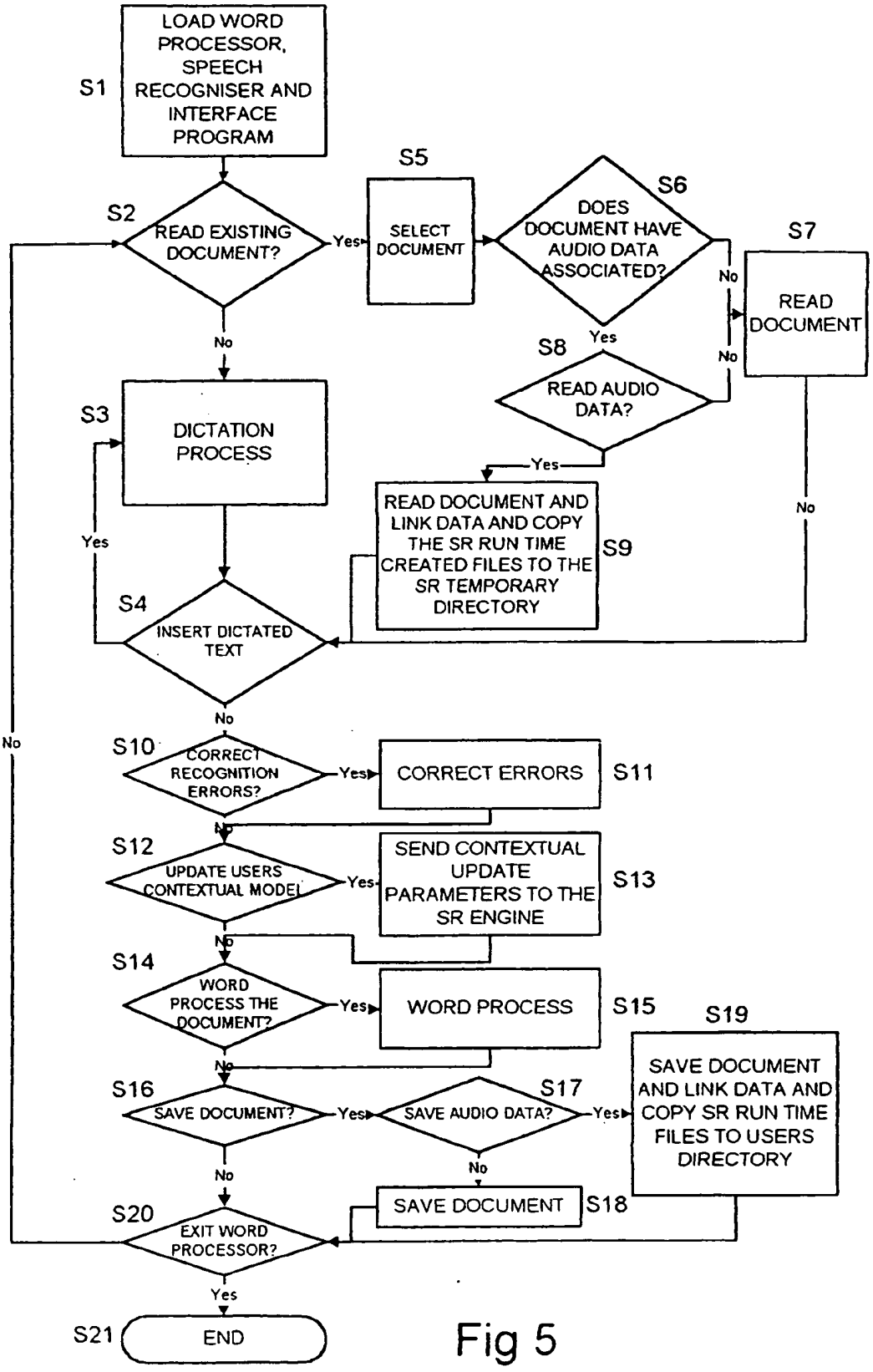

S1
LOAD WORD PROCESSOR, SPEECH RECOGNISER AND INTERFACE PROGRAM
S2
READ EXISTING DOCUMENT?
Yes
No
S5
SELECT DOCUMENT
S6
DOES DOCUMENT HAVE AUDIO DATA ASSOCIATED?
Yes
No
S7
READ DOCUMENT
No
S8
READ AUDIO DATA?
Yes
No
S3
DICTATION PROCESS
Yes
READ DOCUMENT AND LINK DATA AND COPY THE SR RUN TIME CREATED FILES TO THE SR TEMPORARY DIRECTORY
S9
S4
INSERT DICTATED TEXT
No
No
S10
CORRECT RECOGNITION ERRORS?
Yes
CORRECT ERRORS
S11
No
S12
UPDATE USERS CONTEXTUAL MODEL
Yes
SEND CONTEXTUAL UPDATE PARAMETERS TO THE SR ENGINE
S13
No
S14
WORD PROCESS THE DOCUMENT?
Yes
WORD PROCESS
S15
S19
No
S16
SAVE DOCUMENT?
Yes
S17
SAVE AUDIO DATA?
Yes
SAVE DOCUMENT AND LINK DATA AND COPY SR RUN TIME FILES TO USERS DIRECTORY
No
No
SAVE DOCUMENT
S18
S20
EXIT WORD PROCESSOR?
Yes
S21
END

Fig 5

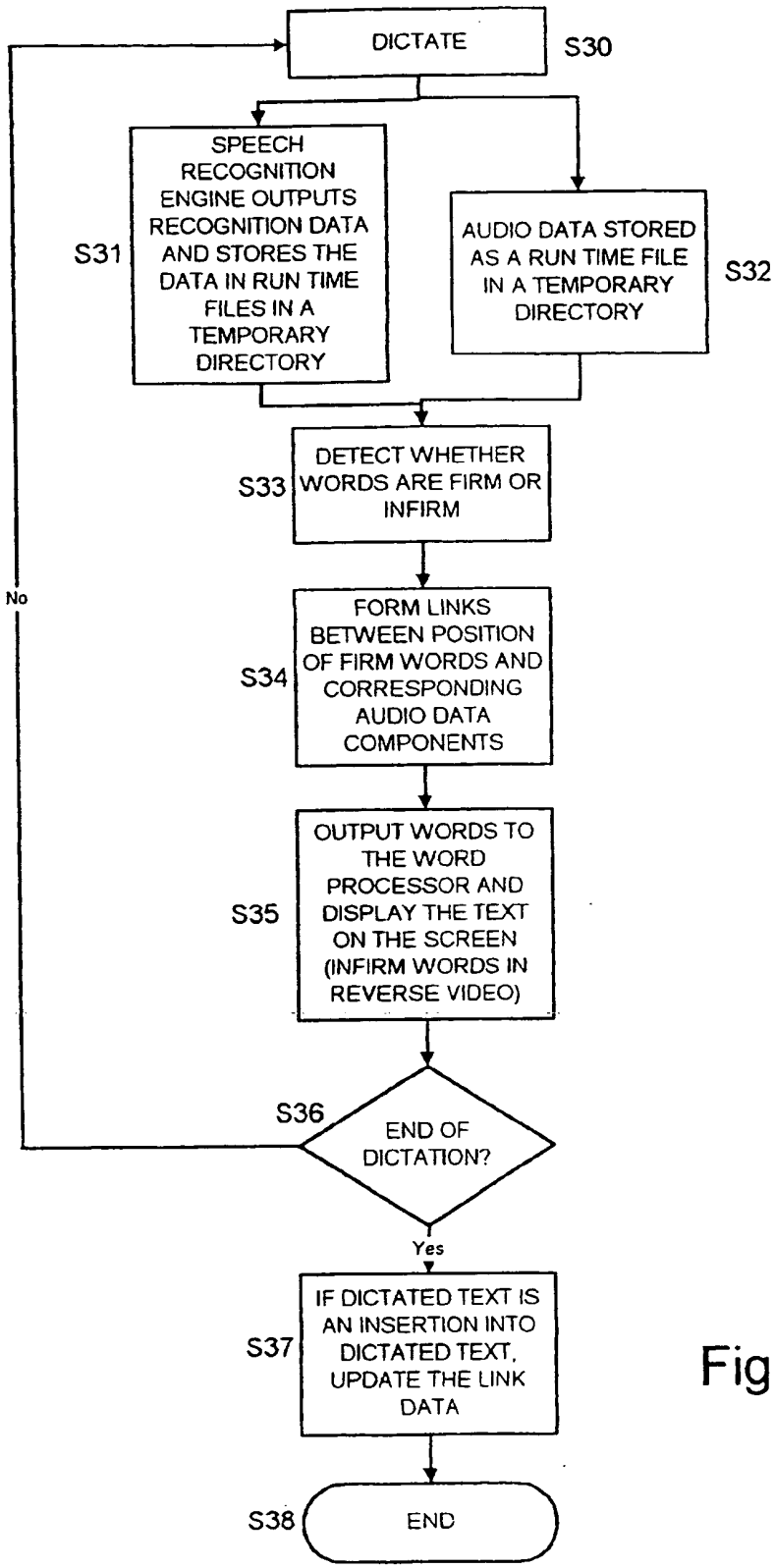

DICTATE
S30
SPEECH RECOGNITION ENGINE OUTPUTS RECOGNITION DATA AND STORES THE DATA IN RUN TIME FILES IN A TEMPORARY DIRECTORY
S31
AUDIO DATA STORED AS A RUN TIME FILE IN A TEMPORARY DIRECTORY
S32
DETECT WHETHER WORDS ARE FIRM OR INFIRM
S33
FORM LINKS BETWEEN POSITION OF FIRM WORDS AND CORRESPONDING AUDIO DATA COMPONENTS
S34
OUTPUT WORDS TO THE WORD PROCESSOR AND DISPLAY THE TEXT ON THE SCREEN (INFIRM WORDS IN REVERSE VIDEO)
S35
S36
END OF DICTATION?
No
Yes
IF DICTATED TEXT IS AN INSERTION INTO DICTATED TEXT, UPDATE THE LINK DATA
S37
END
S38

Fig 6

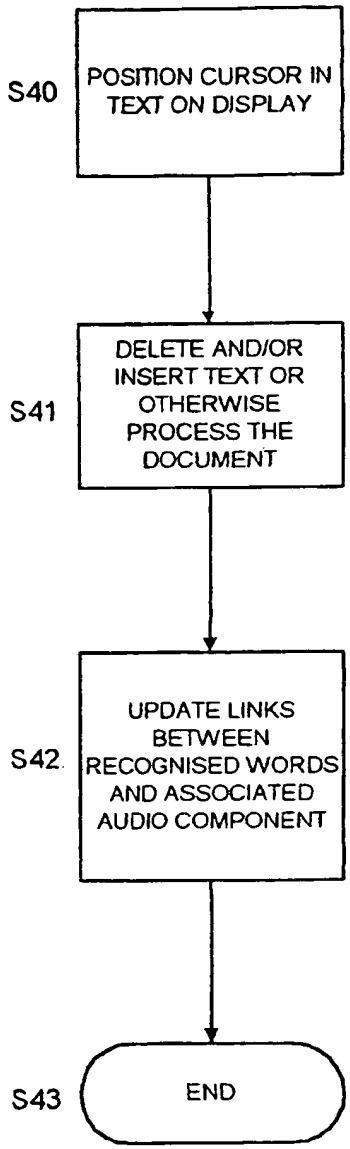

S40
POSITION CURSOR IN TEXT ON DISPLAY
S41
DELETE AND/OR INSERT TEXT OR OTHERWISE PROCESS THE DOCUMENT
S42
UPDATE LINKS BETWEEN RECOGNISED WORDS AND ASSOCIATED AUDIO COMPONENT
S43
END

Fig 7

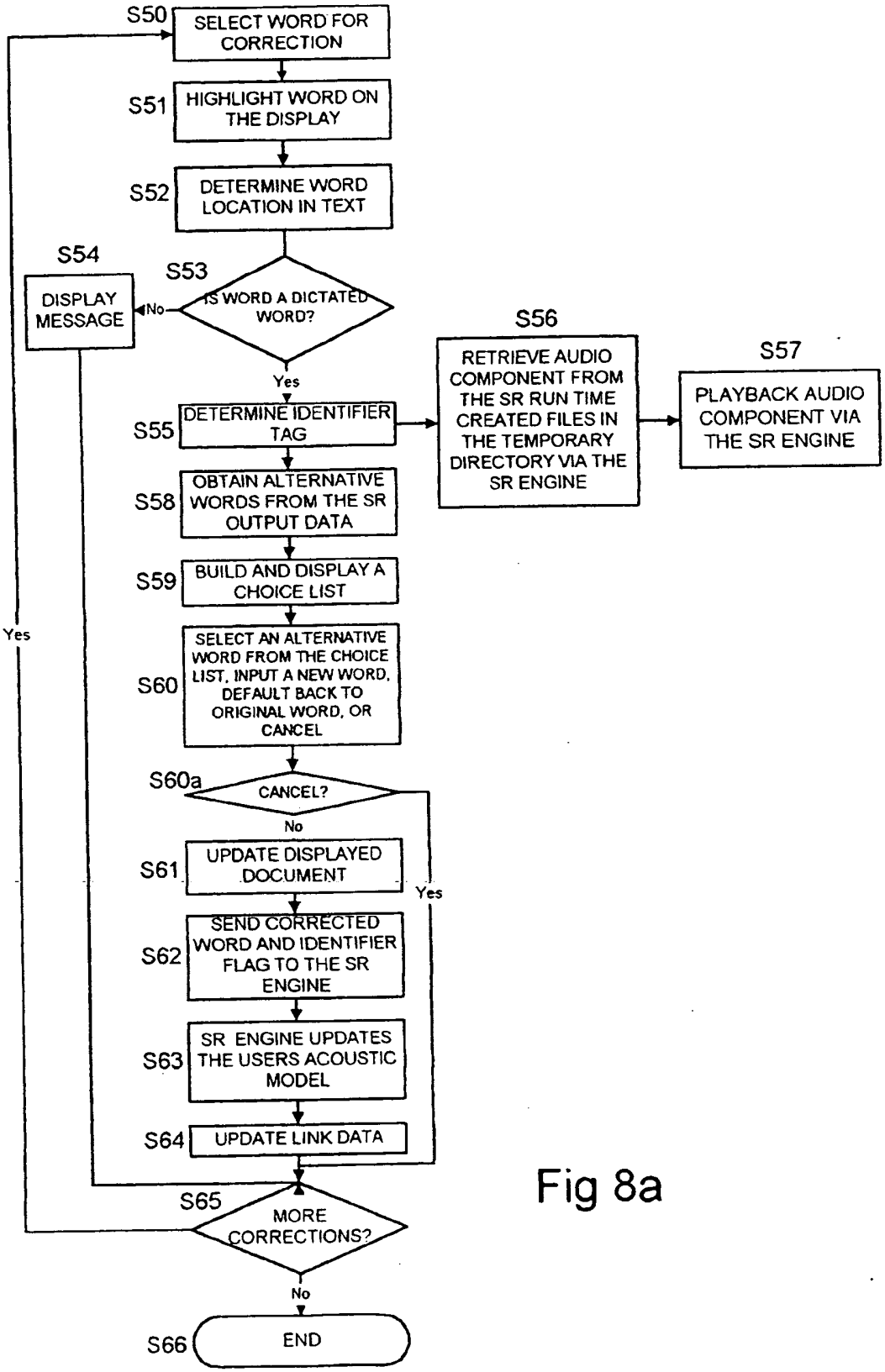

S50
SELECT WORD FOR CORRECTION
S51
HIGHLIGHT WORD ON THE DISPLAY
S52
DETERMINE WORD LOCATION IN TEXT
S53
IS WORD A DICTATED WORD?
No
S54
DISPLAY MESSAGE
Yes
S55
DETERMINE IDENTIFIER TAG
S56
RETRIEVE AUDIO COMPONENT FROM THE SR RUN TIME CREATED FILES IN THE TEMPORARY DIRECTORY VIA THE SR ENGINE
S57
PLAYBACK AUDIO COMPONENT VIA THE SR ENGINE
S58
OBTAIN ALTERNATIVE WORDS FROM THE SR OUTPUT DATA
S59
BUILD AND DISPLAY A CHOICE LIST
S60
SELECT AN ALTERNATIVE WORD FROM THE CHOICE LIST, INPUT A NEW WORD, DEFAULT BACK TO ORIGINAL WORD, OR CANCEL
S60a
CANCEL?
No
Yes
S61
UPDATE DISPLAYED DOCUMENT
S62
SEND CORRECTED WORD AND IDENTIFIER FLAG TO THE SR ENGINE
S63
SR ENGINE UPDATES THE USERS ACOUSTIC MODEL
S64
UPDATE LINK DATA
S65
MORE CORRECTIONS?
Yes
No
S66
END

Fig 8a

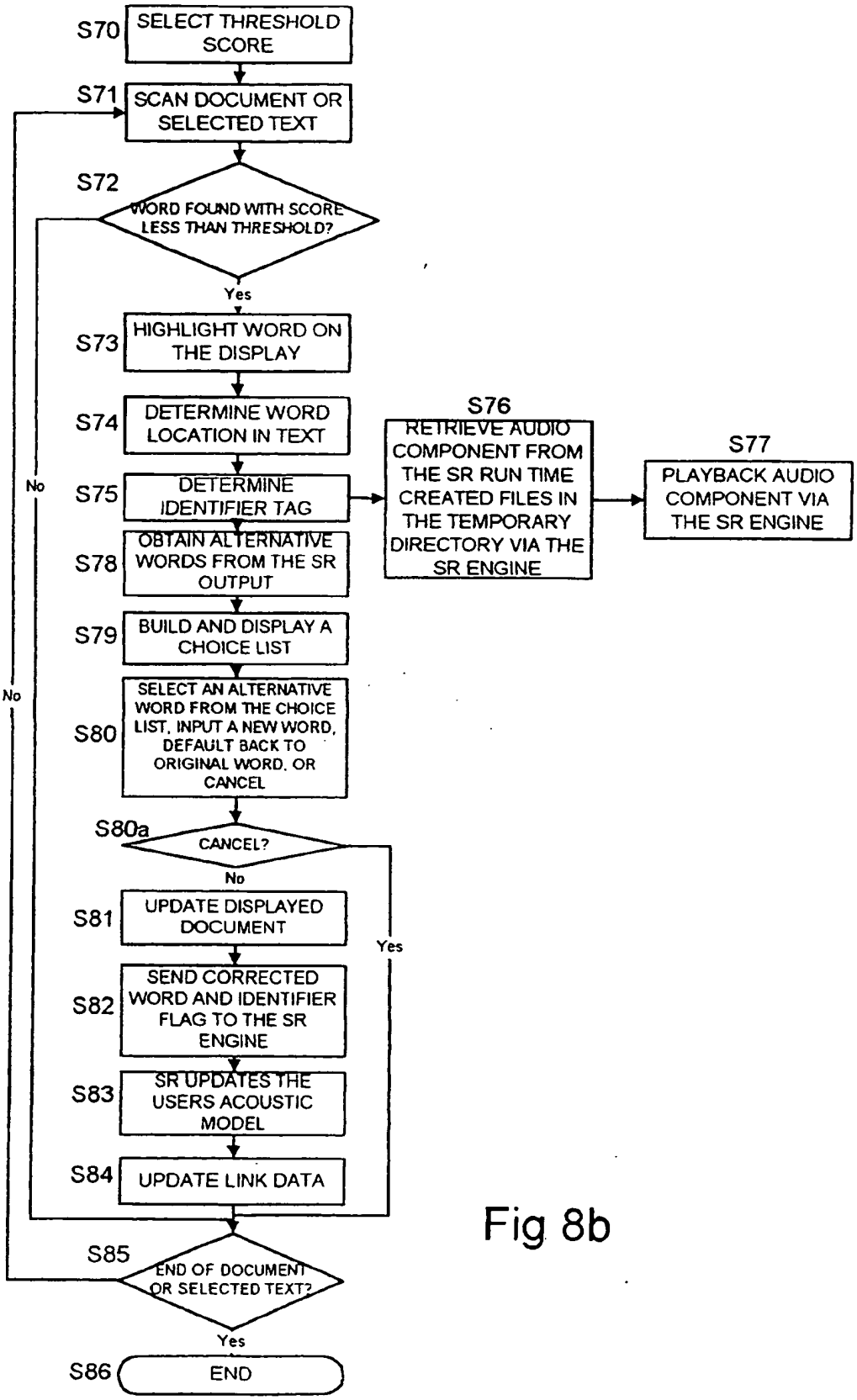

S70
SELECT THRESHOLD SCORE
S71
SCAN DOCUMENT OR SELECTED TEXT
S72
WORD FOUND WITH SCORE LESS THAN THRESHOLD?
Yes
No
S73
HIGHLIGHT WORD ON THE DISPLAY
S74
DETERMINE WORD LOCATION IN TEXT
S75
DETERMINE IDENTIFIER TAG
S76
RETRIEVE AUDIO COMPONENT FROM THE SR RUN TIME CREATED FILES IN THE TEMPORARY DIRECTORY VIA THE SR ENGINE
S77
PLAYBACK AUDIO COMPONENT VIA THE SR ENGINE
S78
OBTAIN ALTERNATIVE WORDS FROM THE SR OUTPUT
S79
BUILD AND DISPLAY A CHOICE LIST
S80
SELECT AN ALTERNATIVE WORD FROM THE CHOICE LIST, INPUT A NEW WORD, DEFAULT BACK TO ORIGINAL WORD, OR CANCEL
S80a
CANCEL?
No
Yes
S81
UPDATE DISPLAYED DOCUMENT
S82
SEND CORRECTED WORD AND IDENTIFIER FLAG TO THE SR ENGINE
S83
SR UPDATES THE USERS ACOUSTIC MODEL
S84
UPDATE LINK DATA
S85
END OF DOCUMENT OR SELECTED TEXT?
No
Yes
S86
END

Fig 8b

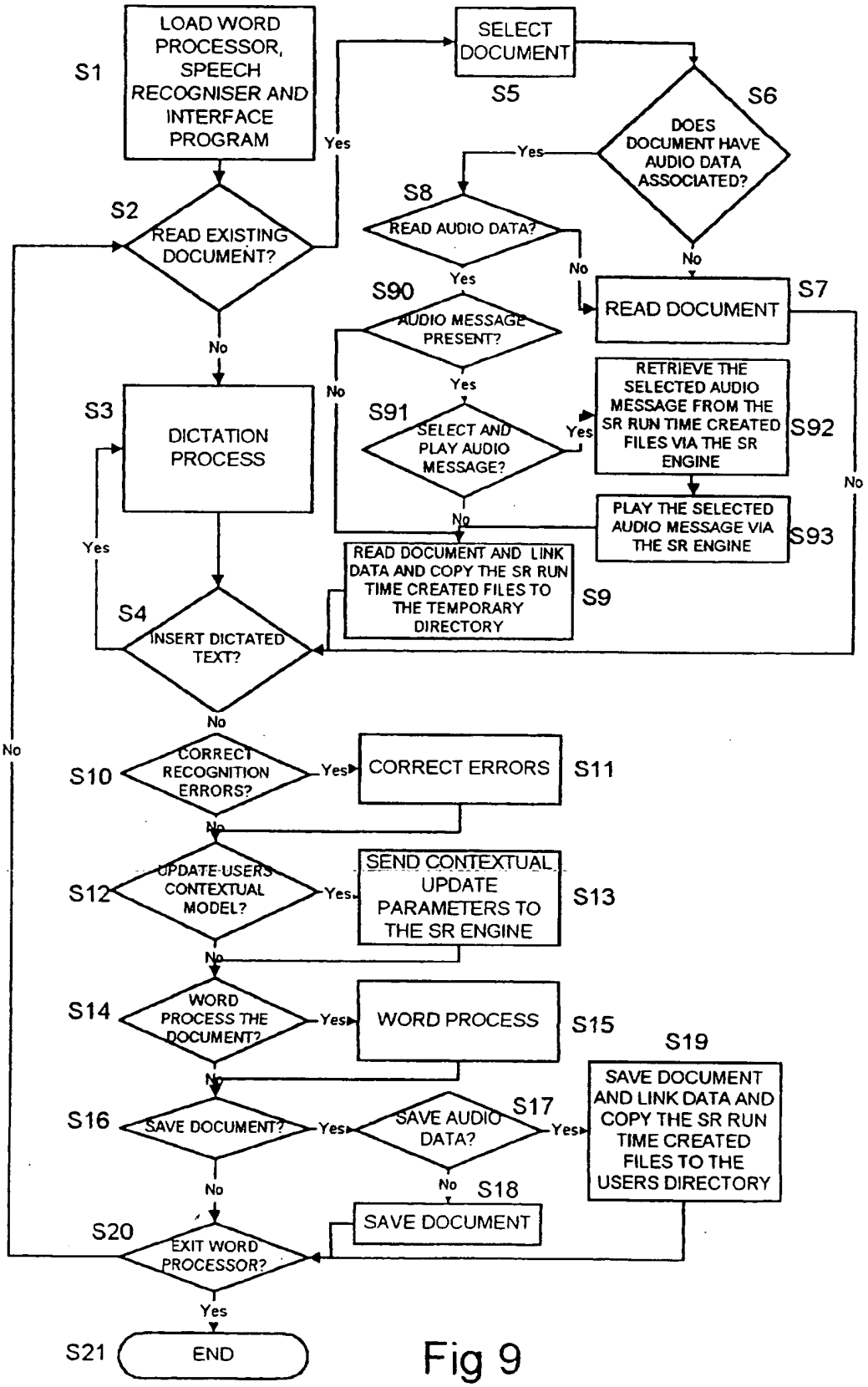

S1
LOAD WORD PROCESSOR, SPEECH RECOGNISER AND INTERFACE PROGRAM
S2
READ EXISTING DOCUMENT?
Yes
No
S3
DICTATION PROCESS
S4
INSERT DICTATED TEXT?
Yes
No
S5
SELECT DOCUMENT
S6
DOES DOCUMENT HAVE AUDIO DATA ASSOCIATED?
Yes
No
S7
READ DOCUMENT
S8
READ AUDIO DATA?
Yes
No
S90
AUDIO MESSAGE PRESENT?
Yes
No
S91
SELECT AND PLAY AUDIO MESSAGE?
Yes
No
S92
RETRIEVE THE SELECTED AUDIO MESSAGE FROM THE SR RUN TIME CREATED FILES VIA THE SR ENGINE
S93
PLAY THE SELECTED AUDIO MESSAGE VIA THE SR ENGINE
S9
READ DOCUMENT AND LINK DATA AND COPY THE SR RUN TIME CREATED FILES TO THE TEMPORARY DIRECTORY
S10
CORRECT RECOGNITION ERRORS?
Yes
No
S11
CORRECT ERRORS
S12
UPDATE USERS CONTEXTUAL MODEL?
Yes
No
S13
SEND CONTEXTUAL UPDATE PARAMETERS TO THE SR ENGINE
S14
WORD PROCESS THE DOCUMENT?
Yes
No
S15
WORD PROCESS
S16
SAVE DOCUMENT?
Yes
No
S17
SAVE AUDIO DATA?
Yes
No
S18
SAVE DOCUMENT
S19
SAVE DOCUMENT AND LINK DATA AND COPY THE SR RUN TIME CREATED FILES TO THE USERS DIRECTORY
S20
EXIT WORD PROCESSOR?
Yes
No
S21
END

Fig 9

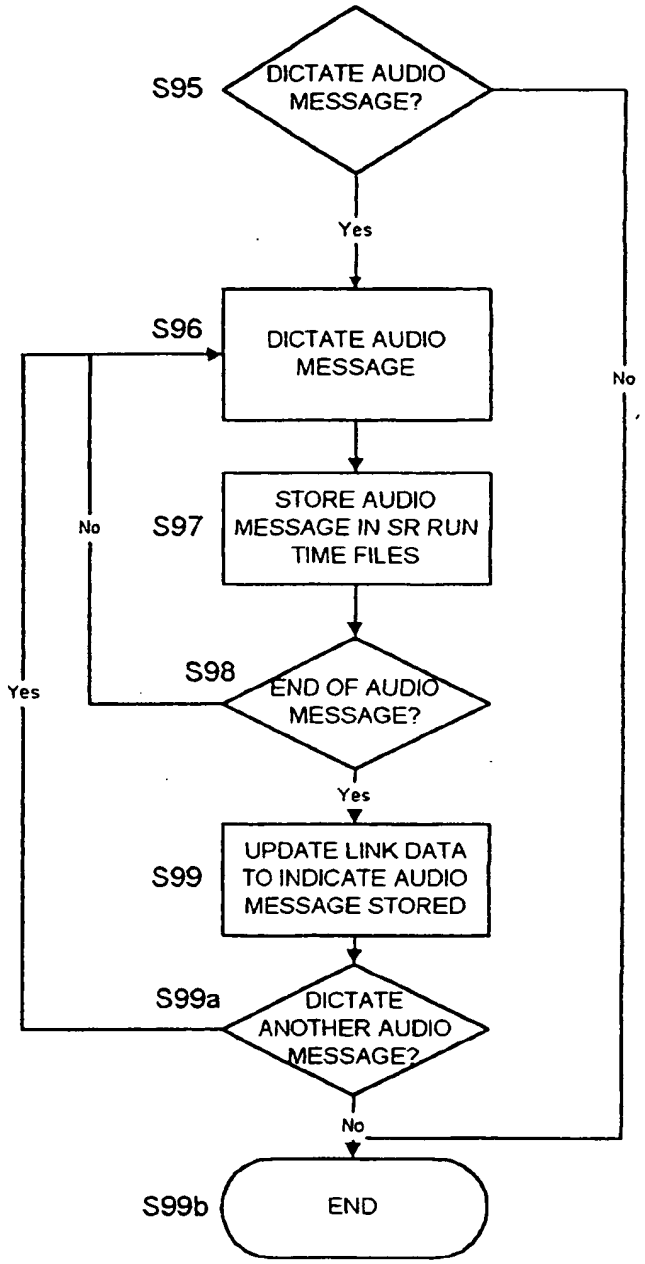

S95
DICTATE AUDIO MESSAGE?
Yes
No
S96
DICTATE AUDIO MESSAGE
S97
STORE AUDIO MESSAGE IN SR RUN TIME FILES
No
S98
END OF AUDIO MESSAGE?
Yes
Yes
S99
UPDATE LINK DATA TO INDICATE AUDIO MESSAGE STORED
S99a
DICTATE ANOTHER AUDIO MESSAGE?
No
S99b
END

Fig 10

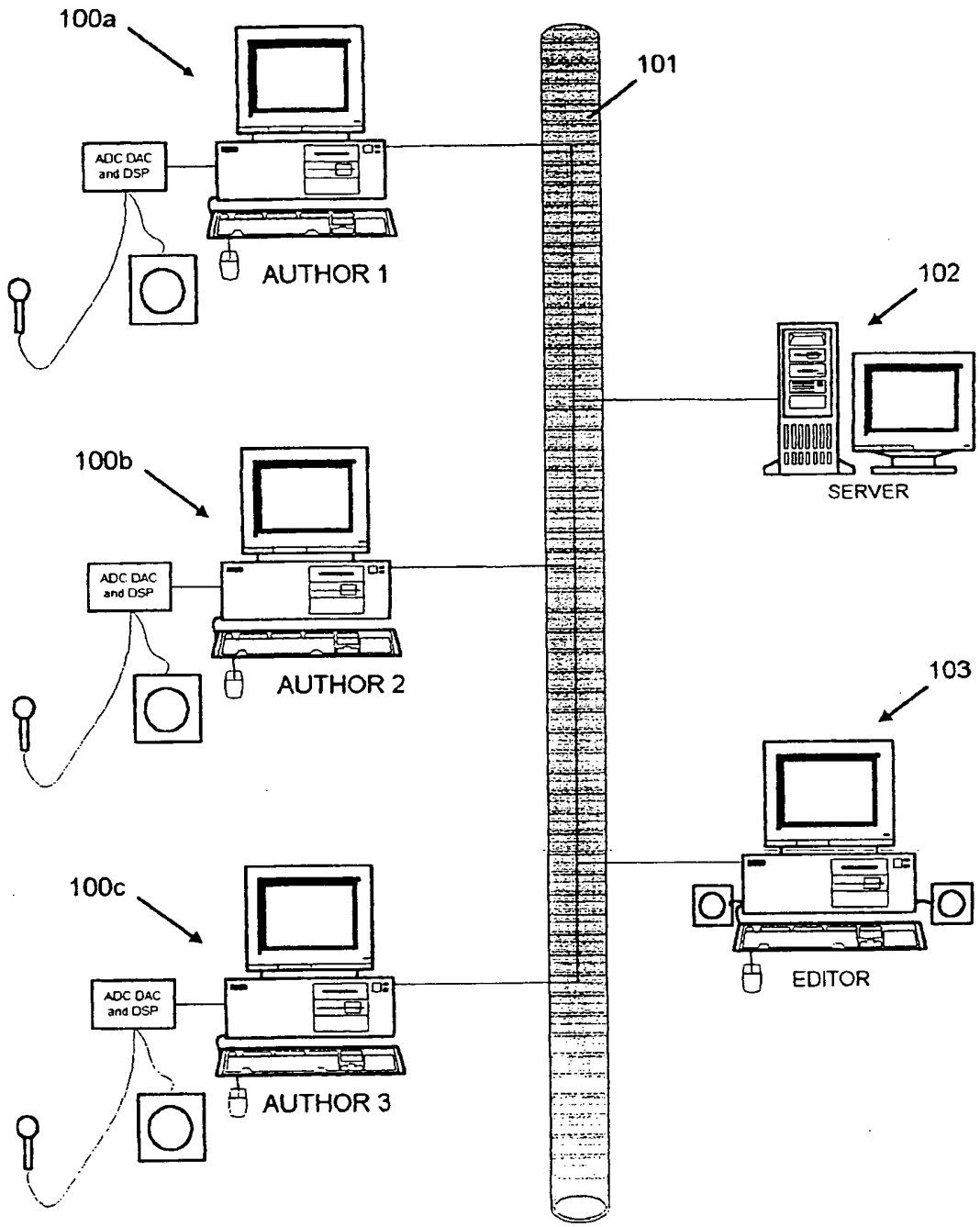

100a
101
ADC DAC and DSP
AUTHOR 1
102
SERVER
100b
ADC DAC and DSP
AUTHOR 2
103
100c
ADC DAC and DSP
AUTHOR 3
EDITOR

Fig 11

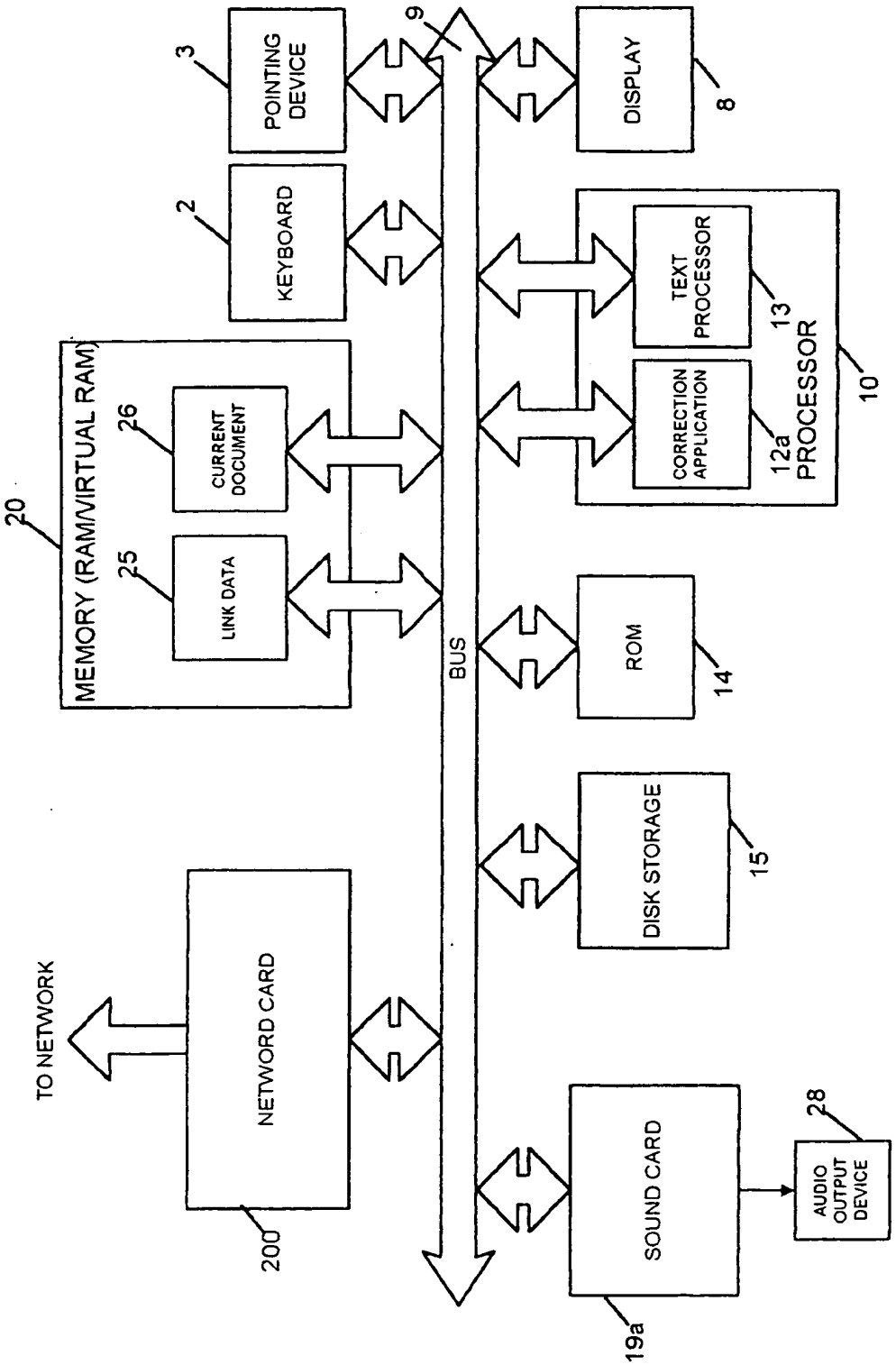
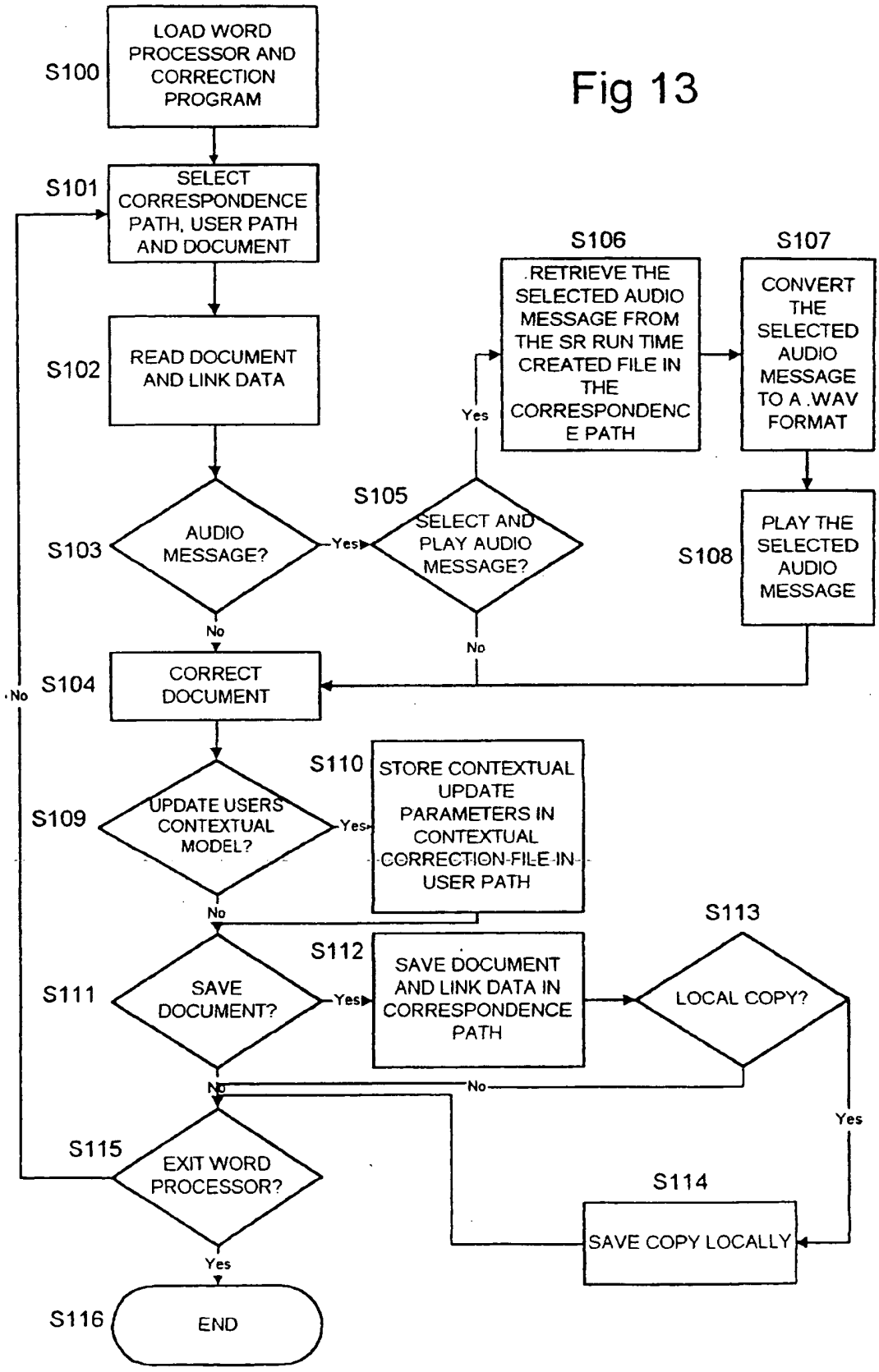

POINTING DEVICE
3
KEYBOARD
2
9
DISPLAY
8
TEXT PROCESSOR
13
CORRECTION APPLICATION
12a
PROCESSOR
10
MEMORY (RAM/VIRTUAL RAM)
20
CURRENT DOCUMENT
26
LINK DATA
25
BUS
ROM
14
DISK STORAGE
15
TO NETWORK
NETWORD CARD
200
SOUND CARD
19a
AUDIO OUTPUT DEVICE
28

Fig 12

Fig 13


S100
LOAD WORD PROCESSOR AND CORRECTION PROGRAM
S101
SELECT CORRESPONDENCE PATH, USER PATH AND DOCUMENT
S102
READ DOCUMENT AND LINK DATA
S103
AUDIO MESSAGE?
Yes
No
S105
SELECT AND PLAY AUDIO MESSAGE?
Yes
No
S106
RETRIEVE THE SELECTED AUDIO MESSAGE FROM THE SR RUN TIME CREATED FILE IN THE CORRESPONDENCE PATH
S107
CONVERT THE SELECTED AUDIO MESSAGE TO A .WAV FORMAT
S108
PLAY THE SELECTED AUDIO MESSAGE
S104
CORRECT DOCUMENT
No
S109
UPDATE USERS CONTEXTUAL MODEL?
Yes
No
S110
STORE CONTEXTUAL UPDATE PARAMETERS IN CONTEXTUAL CORRECTION FILE IN USER PATH
S111
SAVE DOCUMENT?
Yes
No
S112
SAVE DOCUMENT AND LINK DATA IN CORRESPONDENCE PATH
S113
LOCAL COPY?
No
Yes
S114
SAVE COPY LOCALLY
S115
EXIT WORD PROCESSOR?
No
Yes
S116
END

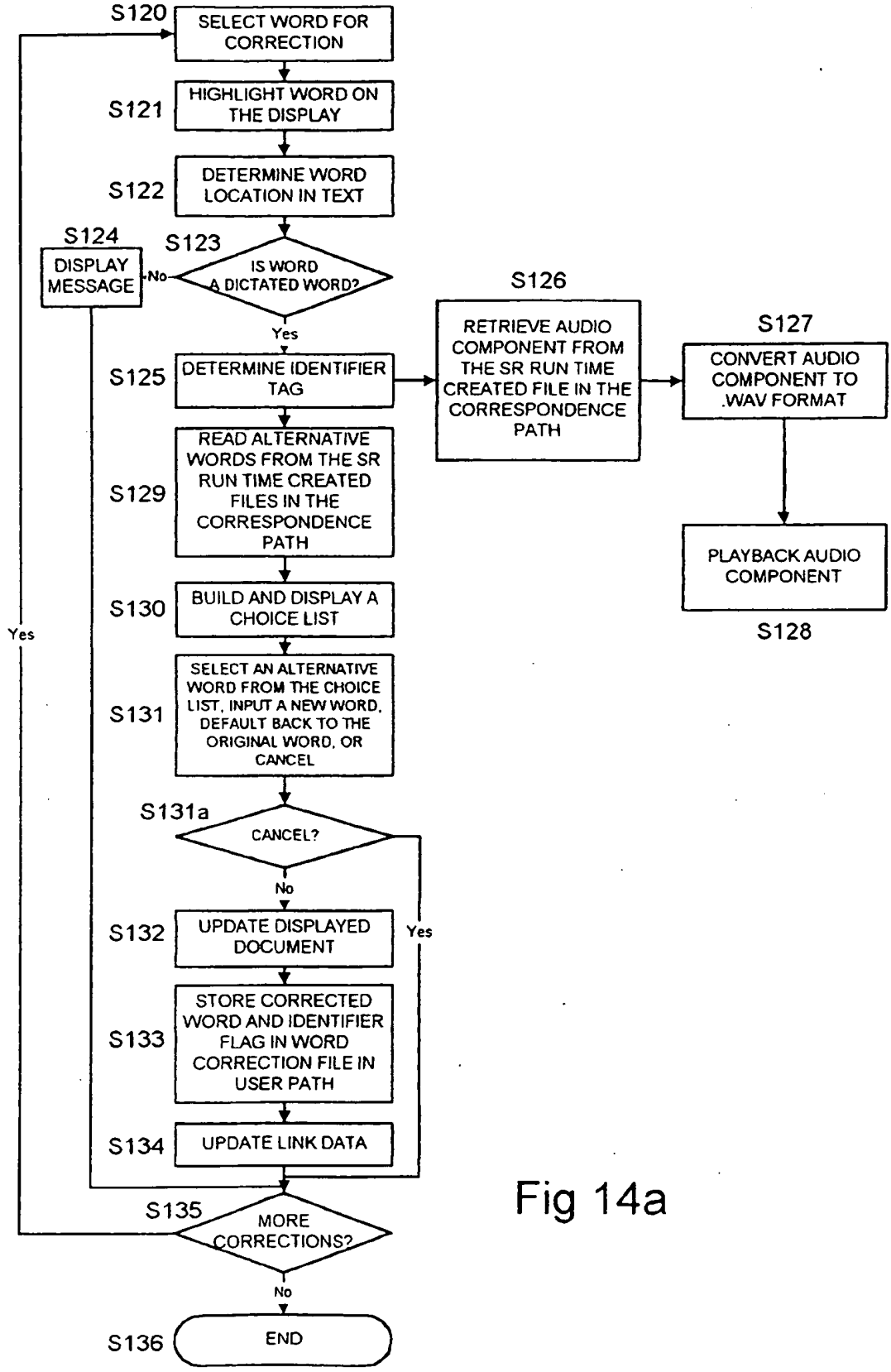

S120
SELECT WORD FOR CORRECTION
S121
HIGHLIGHT WORD ON THE DISPLAY
S122
DETERMINE WORD LOCATION IN TEXT
S123
IS WORD A DICTATED WORD?
No
S124
DISPLAY MESSAGE
Yes
S125
DETERMINE IDENTIFIER TAG
S126
RETRIEVE AUDIO COMPONENT FROM THE SR RUN TIME CREATED FILE IN THE CORRESPONDENCE PATH
S127
CONVERT AUDIO COMPONENT TO .WAV FORMAT
PLAYBACK AUDIO COMPONENT
S128
S129
READ ALTERNATIVE WORDS FROM THE SR RUN TIME CREATED FILES IN THE CORRESPONDENCE PATH
S130
BUILD AND DISPLAY A CHOICE LIST
S131
SELECT AN ALTERNATIVE WORD FROM THE CHOICE LIST, INPUT A NEW WORD, DEFAULT BACK TO THE ORIGINAL WORD, OR CANCEL
S131a
CANCEL?
Yes
No
S132
UPDATE DISPLAYED DOCUMENT
S133
STORE CORRECTED WORD AND IDENTIFIER FLAG IN WORD CORRECTION FILE IN USER PATH
S134
UPDATE LINK DATA
S135
MORE CORRECTIONS?
Yes
No
S136
END

Fig 14a

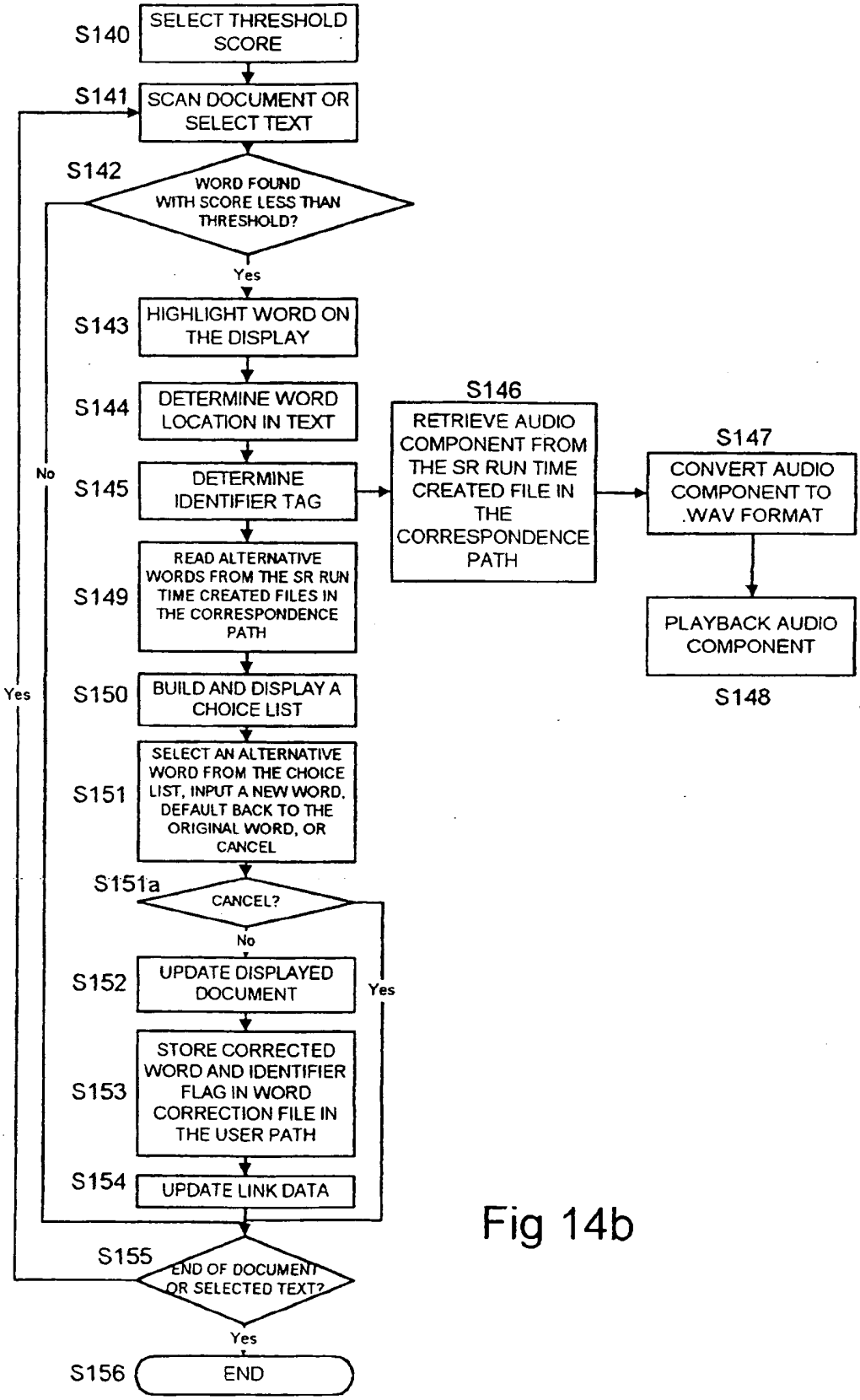

S140
SELECT THRESHOLD SCORE
S141
SCAN DOCUMENT OR SELECT TEXT
S142
WORD FOUND WITH SCORE LESS THAN THRESHOLD?
Yes
No
S143
HIGHLIGHT WORD ON THE DISPLAY
S144
DETERMINE WORD LOCATION IN TEXT
S145
DETERMINE IDENTIFIER TAG
S146
RETRIEVE AUDIO COMPONENT FROM THE SR RUN TIME CREATED FILE IN THE CORRESPONDENCE PATH
S147
CONVERT AUDIO COMPONENT TO .WAV FORMAT
PLAYBACK AUDIO COMPONENT
S148
S149
READ ALTERNATIVE WORDS FROM THE SR RUN TIME CREATED FILES IN THE CORRESPONDENCE PATH
S150
BUILD AND DISPLAY A CHOICE LIST
Yes
S151
SELECT AN ALTERNATIVE WORD FROM THE CHOICE LIST, INPUT A NEW WORD, DEFAULT BACK TO THE ORIGINAL WORD, OR CANCEL
S151a
CANCEL?
No
Yes
S152
UPDATE DISPLAYED DOCUMENT
S153
STORE CORRECTED WORD AND IDENTIFIER FLAG IN WORD CORRECTION FILE IN THE USER PATH
S154
UPDATE LINK DATA
S155
END OF DOCUMENT OR SELECTED TEXT?
Yes
S156
END

Fig 14b

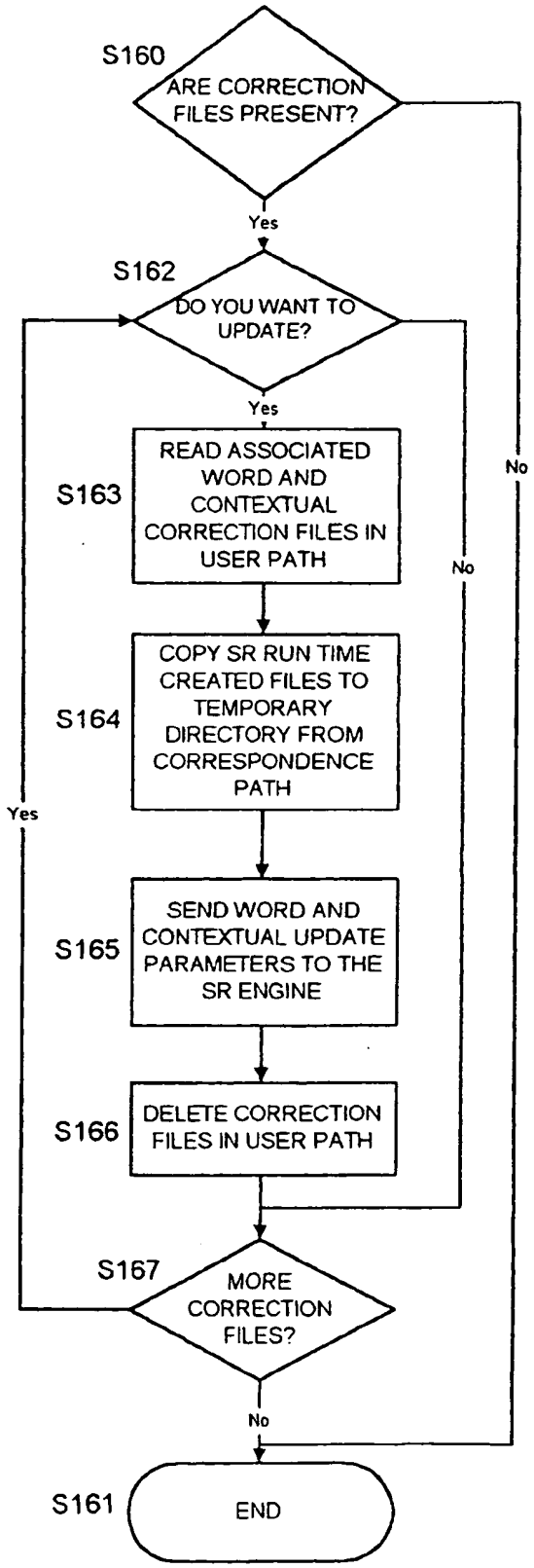

S160
ARE CORRECTION FILES PRESENT?
Yes
No
S162
DO YOU WANT TO UPDATE?
Yes
No
S163
READ ASSOCIATED WORD AND CONTEXTUAL CORRECTION FILES IN USER PATH
S164
COPY SR RUN TIME CREATED FILES TO TEMPORARY DIRECTORY FROM CORRESPONDENCE PATH
S165
SEND WORD AND CONTEXTUAL UPDATE PARAMETERS TO THE SR ENGINE
S166
DELETE CORRECTION FILES IN USER PATH
S167
MORE CORRECTION FILES?
Yes
No
S161
END

Fig15

DATA PROCESSING METHOD AND APPARATUS

The present invention generally relates to the field of data processing and in particular the field of processing the output of a speech recognition engine.

The use of speech recognition as an alternative method of inputting data to a computer is becoming more prevalent as speech recognition algorithms become ever more sophisticated and the processing capabilities of modern computers increases. Speech recognition systems are particularly attractive for people wishing to use computers who do not have keyboard skills.

There are several speech recognition systems currently on the market which can operate on a desktop computer. One such system is called DragonDictate (Trade Mark). This system allows a user to input both speech data and speech commands. The system can interface with many different applications to allow the recognised text output to be directly input into the application, e.g. a word processor. This system, however, suffers from the disadvantage that there is no audio recording of the dictation stored which can be replayed to aid the correction of the recognised text.

Another system which is currently on the market is IBM VoiceType version 1.1 (Trade Mark). In this system the recognised text from the speech recognition engine is input directly into a proprietary text processor and

audio data is stored. This system, however, does not allow the recognised text to be directly input into any other application. The dictated text can only be input directly into the proprietary text processor provided whereupon at the end of dictation the text can be cut and pasted into other applications. Corrections to the dictated text in order to update the speech recogniser models can only be carried out within the text processor window. Text for recognition correction can be selected and the audio component corresponding to the text is played back to assist in the correction process. When all of the corrections have been completed, the text can either be saved or cut ready for pasting into another application. Either of these operations can cause the corrections made to be used to update the speech recogniser: the user has limited control over when the updates are made.

Not only is this system disadvantaged in not allowing direct dictation into applications, the system also does not allow the audio data to be stored in association with the text when the document is saved or when the text is cut and pasted into another application. Even a simple text processing operation, e.g. an insertion operation within a body of text, will prevent the playback of the audio component for that body of text including the change.

It is an object of one aspect of the present

invention to provide an interface between the output of a speech recognition engine and an application capable of processing the output which operates in a data processing apparatus to link the relationship between the output data and the audio data to allow the audio data to be played back for any output data which has been dictated even if the data as a whole has been processed in such a way as to move, reorder, delete, insert or format the data.

According to one aspect the present invention provides a data processing apparatus comprising input means for receiving recognition data and corresponding audio data from a speech recognition engine, the recognition data including a string of recognised data characters and audio identifiers identifying audio components corresponding to a character component of the recognised characters; processing means for receiving and processing the input recognised characters to replace, insert, and/or move characters in the recognised characters and/or to position the recognised characters; link means for forming link data linking the audio and identifiers to the characters component positions in the character string even after processing; display means for displaying the characters being processed by the processing means; user operable selection means for selecting characters in the displayed characters for audio playback, where the link data identifies any

selected audio components, if present, which are linked to the selected characters; and audio playback means for playing back the selected audio components in the order of the character component positions in the character string.

Thus, in accordance with this aspect of the present invention, positional changes of characters in the character string due to processing operations are monitored and the links which identify the corresponding audio component are updated accordingly. In this way, the corresponding audio component for any dictated character in the character string can be immediately identified even after processing. This allows for the audio component associated with any character to be played back by a selection operation by a user. This feature greatly enhances the ability to correct incorrectly recognised characters since a user will be able to hear what was dictated in order to decide what was actually said rather than what the speech recogniser recognised. This feature of being able to play back audio components corresponding to the characters is maintained even when dictated characters are inserted into previously dictated characters.

In the present invention the character data output from the speech recognition engine can comprise text or symbols in any language, numerals or any unicode. The characters can comprise words forming text or any unicode

characters and the system can be configured to recognise dictated numbers and input the corresponding numeric characters to the application instead of the word descriptions.

The processing means of the present invention can comprise any application running on a processor which enables character data from a speech recognition engine to be entered and manipulated, e.g. a word processor, presentation applications such as Microsoft PowerPoint (Trade Mark) spreadsheets such as Excel (Trade Mark), email applications and CAD applications. In this aspect of the present invention the dictated character positions in the document, drawing or product of the application is linked to the corresponding audio component by link data.

In one aspect of the present invention the link data and audio data can all be stored. In this way the audio data is maintained for playback at a later time when, for instance, it may be wished to carry out corrections to correct speech recognition errors. The storage of the character data, link data and the audio data allows for corrections to be postponed or even delegated to another person on another machine.

Corrections to the incorrectly recognised character data can be made by correcting the character string which causes the playback of the audio component. The characters can then be corrected and the corrected

characters and the audio identifier for the audio component corresponding to the corrected characters are passed to the speech recognition engine for updating user models used in the recognition process.

Where the output of the speech recognition engine includes a list of alternative characters together with an indicator which indicates the likelihood that the word is correct, when a word is selected for correction, a choice list can be displayed which comprises the alternative words listed alphabetically for ease of use. Corrections can then be carried out either by selecting one of the alternative characters or entering a new character.

In one embodiment, in order to maintain the links between the character components and the corresponding audio components, a list of character locations in the character string and positions in the corresponding audio components is kept. Where the character string is formed of a plurality of separate dictated passages, the audio data is separately stored and the list identifies in which of the stored audio passages and at which position the audio component lies in the audio passage.

In addition to the updating of the speech recognition model due to incorrectly recognised words, a passage of characters, or all of the characters, can be selected for updating the contextual model used by the speech recognition engine. Thus, in this embodiment of

the invention the operator has control over when the contextual model is to be updated based on the corrections made to the characters.

It is an object of another aspect of the present invention to enable audio messages to be recorded and stored in association with a file containing character data output from a speech recognition engine to allow instructions or a reminder to be recorded.

In accordance with another aspect of the present invention there is provided data processing apparatus comprising means for receiving recognition data from a speech recognition engine and corresponding audio data, the recognition data including recognised characters; display means for displaying the recognised characters; storage means for storing the recognised characters as a file; means for selectively disabling the display and storage of recognised characters or recognition carried out by the speech recognition engine for a period of time; and means for storing the audio data for a period of time in the storage means as an audio message associated with the file.

It is an object of another aspect of the present invention to provide for the automatic detection of possibly incorrectly recognised characters in the character data output from the speech recognition engine.

In accordance with this aspect of the present invention, there is provided data correction apparatus

comprising means for receiving recognition data from a speech recognition engine, said recognition data including recognised characters representing the most likely characters, and a likelihood indicator for each character indicating the likelihood that the character is correct; display means for displaying the recognised character; automatic error detection means for detecting possible errors in recognition of characters in the recognised characters by scanning the likelihood indicators for the recognised characters and detecting if the likelihood indicator for a character is below the likelihood threshold, whereby said display means highlights at least the first, if any, character having a likelihood indicator below the likelihood threshold; user operable selection means for selecting a character to replace an incorrectly recognised character highlighted in the recognised characters; and correction means for replacing the incorrectly recognised character and the selected character to correct the recognised characters.

The likelihood threshold can be selectively set by a user to a suitable level to reduce the number of characters which are falsely identified as incorrectly recognised whilst increasing the chances of correctly identifying incorrectly recognised characters. The provision of automatic detection of possible recognition errors can significantly decrease the time taken for

correcting character data.

Embodiments of the present invention will now be described with reference to the accompanying drawings, in which:

Figure 1 is a schematic drawing of a speech recognition system in accordance with one embodiment of the present invention;

Figure 2 is a schematic diagram of the internal structure of the speech recognition system;

Figure 3 is a table representing the data output from the speech recognition engine;

Figure 4 illustrates the data structure of the link data file;

Figure 5 is a flow chart illustrating the overall operation of the speech recognition system in accordance with one embodiment of the present invention;

Figure 6 is a flow diagram of the dictation process of figure 5;

Figure 7 is a flow diagram of the word processing process of figure 5;

Figure 8a is a flow diagram of the manual correction process of figure 5;

Figure 8b is a flow diagram of the automatic correction process of figure 5;

Figure 9 is a flow diagram of the overall operation of the speech recognition system in accordance with another embodiment of the present invention in which

audio messages can be played;

Figure 10 is a flow diagram of an optional step for the dictation of an audio message in the sequence of figure 9;

Figure 11 is a schematic drawing of a network of speech recognition systems comprising author work stations wherein the network is provided with an editor work station which can access and edit documents in the author work stations;

Figure 12 is a schematic diagram of the internal structure of the editor work station;

Figure 13 is a flow diagram of the overall operation of the operation of the editor work station of figure 11;

Figure 14a is a flow diagram of the manual correction process of figure 13;

Figure 14b is a flow diagram of the automatic correction process of figure 13; and

Figure 15 is a flow diagram of the speech recognition model update process which is carried out by the author work stations after corrections have been made to recognised text by the editor work station.

A specific embodiment will now be described with application to word processing of text output of a speech recognition engine.

Referring to figure 1 there is illustrated a speech recognition system in accordance with one embodiment of the present invention which comprises an IBM (Trade Mark)

compatible PC (personal computer) 1 having a keyboard 2 for inputting and correcting text and a pointing device 3 which in this embodiment is a mouse. Software applications are loaded into the computer from computer storage medium such as the floppy disc 4, an optical disk (CD ROM), or digital tape. The software applications comprise the speech recognition application which comprises the speech recognition engine, the application for processing text such as a word processor and the interface application to control the flow of text into the text processing application, to control the flow of updating information from the text processing application to the speech recognition application and for maintaining links between the text and the audio data.

The system is provided with a microphone 5, a loudspeaker 6 and an interface device 7. During dictation the audio signal from the microphone 5 is input into the interface device 7 which includes an analog to digital converter and a digital signal processor to digitise and condition the signal for input into the computer 1. During playback of the recorded audio signal, the audio signal is output from the computer 1 to the interface device 7 in digital form and is converted to an analog signal by a digital to analog converter within the interface device 7. The analog signal is then output from the interface device 7 to play back the audio recording.

In the specific embodiment of the present invention the interface device 7 is provided with the IBM VoiceType system. Also, the speech recognition engine used in the specific example is the IBM VoiceType speech recognition engine. The present invention is not, however, limited to any specific speech recognition engine and can also be used with any conventional hardware for recording and playing back sound in a personal computer, e.g. in an IBM compatible machine the sixteen bit sound blaster compatible standard can be used. The present invention can be used with either continuous or discrete speech recognition engines.

Referring now to figure 2, this diagram illustrates a schematic overview of the internal architecture of the computer. A bus 9 links all of the components of the system and the Read Only Memory (ROM) 14 containing conventional systems programs and data. The processor 10 runs three applications simultaneously: the speech recognition engine application 11, the speech recognition interface application 12 and the text processor application 13. The memory 20, which can comprise random access memory (RAM) or in a Windows (Trade Mark) environment, virtual RAM. Within the memory 20 data is stored for the speech recognition engine application 11. This data comprises a user model 21 which can be updated to improve the accuracy of the recognition, a language model 22 and a dictionary 23 to which a user can add new

words. The user model 21 comprises an acoustic model and a contextual model. During operation of the speech recognition engine application 11 the application utilises the user model 21, the language model 22 and the dictionary 23 in the memory 20 and outputs speech recognition data 24 to the memory 20. The speech recognition interface application 12 receives the speech recognition output data 24 and forms link data 25. The text component of the speech recognition output data 24 is also passed by the speech recognition interface application 12 to the text processor application 13 to form a current document 26 in the memory. The display 8 displays the text of the current document 26 stored in the memory 20 and the keyboard 2 can be used to insert, delete and move text. The pointing device 3 can also be used to select text and word processing operations in the conventional well known manner within Windows applications.

The system is also provided with non-volatile storage in the form of disk storage 15. Within the disk storage 15 two directories are provided. A temporary directory used by the speech recognition engine 11 for the storage of run time files which contain the speech recognition output data. A user's directory is also provided for the storage of document files by the text processor application 13 and associated link data formed by the speech recognition interface 12.

An audio input device 16 inputs the dictated audio signal to an analog to digital converter 17. Although in figure 1 the audio input device 16 is illustrated to be a microphone 5, the audio input could alternatively comprise a pre-recorded signal source, e.g. a digital audio tape (DAT). The digitised signal from the analog to digital converter 17 is then passed to a digital signal processor 18 for conditioning of the signal before input to the input/output device 19 of the computer 1. In this way the speech recognition engine application 11 is able to read the digitised input audio data via the bus 9 and output speech recognition output data 24 into the memory 20.

When the speech recognition interface application 12 interacts with the text processor application 13 following the selection of text for audio playback by the user using the pointing device 3, audio data which is stored in the temporary directory in the disc storage 15 is accessed and output over the bus 9 via the input/output device 19 to a digital to analog converter 27 to generate an analog audio signal to drive an audio output device 28 for playback of the audio signal selected by the user.

In the specific embodiment the audio data is stored in one or more files in the temporary directory of the disk storage 15 since the storage audio data requires a great deal of storage capacity and it is impractical to

hold audio data of any length in the volatile memory 20.

In the specific embodiment the operating system operating by the processor 10 is Windows 3.1, 3.11, 95 or NT. The text processor application 13 can be any word processor such as Microsoft Word (Trade Mark), Wordperfect (Trade Mark) or Lotus Word Pro (Trade Mark). The speech recognition engine application 11 is the IBM VoiceType.

When the speech recognition engine application 11 is activated and receives audio data via the interface device 7, the speech recognition output data 24 is temporarily held in the volatile memory 20. The output data is then passed to files which are opened in the temporary directory of the disk storage 15. The audio data for each period of dictation is stored in a single file.

Also in temporary directory on the disc storage 15, two files are stored by the speech recognition engine application 11 which includes the information illustrated in tabular form in figure 3. For each period of dictation an audio data file, and a pair of information files are generated containing the information illustrated in figure 3. Each of the words recognised is identified by an identifier tag which identifies the position in the sequence of word. Also, the audio start point and audio end point of the audio component in the associated audio data file is indicated to enable the

retrieval and playback of the audio component corresponding to the word. For each word, a list of alternative words and their scores is given where n is the score, i.e. the likelihood that the word is correct, and w is the word. The list of alternative words is ordered such that the most likely word appears first. Alternatives, if any, are then listed in order with the word having the highest score first and the word having the lowest score last.

The speech recognition interface application 12 receives the output of the speech recognition engine application 11 and forms link data 25 in the volatile memory 20. Figure 4 illustrates the form of the link data for each recognised word output from the speech recognition engine 11. The speech recognition interface application 12 receives the recognised word at the head of the alternative list shown in figure 3 and outputs the word using the dynamic data exchange (DDE) protocol in the Windows operating system. The position of a word in the text in the text processor application 13 is determined by determining the counter number indicating the position of the first character in the text for the word. This character number is entered under the character number field. The link data 25 also includes information identifying where the audio data can be found in the files in the temporary directory of the disk storage 15. This information is provided in the tag

field. The tag field will not only include the identified tag identifying the position of the audio component for a word within a file, it will also include an identification of which file contains the audio component. The next field is the word score which is an indication of the likelihood that the word has been recognised correctly. The next field is the word length field. This gives the number of characters forming the recognised word. The next field in the link data 25 is the character string forming the actual word and this is followed by the vocabulary length field which is a number indicating the number of characters in the vocabulary description string. The final field is the vocabulary description string which is a string of characters describing the vocabulary in which the word recognised by the speech recognition engine application 11 can be found in the dictionary 23.

Figure 5 is an overview of the operation of the embodiment of the present invention. In step S1 the word processor application 13, the speech recognition engine application 11 and the speech recognition interface application 12 is loaded from a storage medium such as the disk storage 15. The programs can of course be loaded from any computer readable storage medium such as optical discs (CD ROM) or digital tape.

Once the programs are loaded, a user can select whether to read an existing document in step S2. If no

existing document is to be read text can be entered using the dictation process step S3 which will be described in more detail hereinafter. When a passage of dictated text is complete, the dictation process is finished and in step S4 the user can decide whether to insert further dictated text. If further dictated text is to be inserted, the process returns to step S3. If no further dictated text is to be inserted then the dictation process is finished.

If in step S2 after the programs have been loaded a user requests that an existing document be read, in step S5 the document to be read is selected and in step S6 it is determined whether the document selected has audio data associated with it. If there is no audio data associated with it, i.e. it is a conventional word processor document, in step S7 the document is read and the process moves to step S4 which is a point at which the document has been loaded and the user can insert dictated text if desired.

If in step S6 it is determined that the document does have audio data associated with it, the user is given the option to read the audio data in step S8. If the user declines to read the audio data then only the document is read in step S7 and the document will be treated within the word processor as a conventional word processor document. If in step S8 the user selects to read the audio data, in step S9 the document is read

together with the associated link data from the user's directory in the disk storage 15 and the speech recogniser run time created files are copied from the user's directory to the temporary directory in the disk storage 15. The document is thus open in the word processor and in step S4 the user can insert dictated text if desired.

If no more dictated text is to be inserted in step S4, in step S10 the user can decide whether to correct recognition errors in the recognised text. If in step S10 it is decided by the user that they are to correct errors then the process moves to step S11 to correct the errors as will be described hereafter.

Once the recognition errors have been corrected by the user or if the recognition error is not to be corrected by the user, the process moves to step S12 wherein the user can decide whether to update the user's contextual model. This is a second form of correction for the speech recognition process. The user model 21 comprises an acoustic model and a contextual model. The recognition errors corrected in step S11 will correct the acoustic model, i.e. the recognition errors. Once all of the recognition errors have been corrected, the contextual model can be updated in step S13 by selecting the text to be used for the update and sending the number of corrected words together with a list of the corrected words to the speech recognition engine for updating the

contextual model.

In step S14 the user can then decide whether or not to word process the document in the conventional manner. If a document is to be word processed, the word processing operation in step S15 is carried out as will be described in more detail hereinafter. This word processing operation can be carried out at any time after or before the dictation process. The document being formed in the word processor can thus comprise a mixture of conventionally entered text, i.e. via the keyboard or via the insertion of text from elsewhere, and directly dictated text.

When the user has finished dictating, inserting and editing the text, in step S16 the user has the option of whether or not to save the document. If the document is to be saved, in step S17 the user is given the option of saving the document without the audio data as a conventional word processor document in step S18, or saving the document together with the link data and audio data in step S19. In step S19, in order to save the link data and audio data, the document and link data, by default, is saved in the user's directory and a copy of the speech recogniser run time created files is made in the user's directory.

Once the document has been saved, the user has the option to exit the word processor in step S20. If the word processor is exited in step S20 the process

terminates in step S21, otherwise the user has the option of whether or not to read an existing document in step S2.

Referring now to figure 6, this document illustrates the dictation process, step S3, of figure 5 in more detail.

In step S30 the dictation is started and in step S31 the speech recognition engine application 11 outputs speech recognition data 24 and stores the data in run time files in a temporary directory of the disk storage 15. Also, the audio data is stored in parallel as a run time file in the temporary directory in step S32. The speech recognition interface application 12 detects whether the most likely words output from the speech recognition engine application 11 are firm or infirm, i.e. whether the speech recognition engine application 11 has finished recognising that word or not in step S33. If the speech recognition engine application 11 has not finished recognising that word, a word is still output as the most likely, but this could change, e.g. when contextual information is taken into consideration. In step S34, the speech recognition interface application 12 forms links between positions of firm words and corresponding audio data components thus forming the link data 25. In step S35 the speech recognition interface application 12 outputs the words to the word processor application 13 and the text is displayed on the screen

with the infirm words being displayed in reverse video format. In step S36 the process determines whether dictation is finished and if has not it returns to step S30. If dictation has finished, in step S37 it is determined whether the dictated text is inserted into previously dictated text and, if so, the link data is updated to take into consideration the change in character positions of the previously dictated words. The dictation process is then terminated in step S38.

Referring now to figure 7, this illustrates the word processing process of step S15 of figure 5 in more detail. In step S40 a user can position the cursor in the text on the display using the keyboard 2 or the pointing device 3. In step S41 the user can delete and/or insert text by, for example, typing using a keyboard or inserting text from elsewhere using conventional word processing techniques. In step S42 the speech recognition interface application 12 updates the links between the recognised words and associated audio components, i.e. the character number in the first field of the link data 25 is amended to indicate the correct character position of the word in the text. The word processing process is then terminated in step S43.

Referring now to figure 8a, this diagram illustrates a manual method of carrying out the error correction of step S11 of figure 5. In step S50 the user selects a word which is believed to be incorrectly recognised for

correction. The selected word is then highlighted on the display in step S51 and in step S52 the speech recognition interface application 12 determines the word location in the text. In step S53 it is determined whether the word is a dictated word or not by comparing the word location with the link data 25. If the word is not a dictated word a message is displayed informing the user that the word is not a dictated word in step S54 and in step S65 the system waits for more corrections. If the word is a dictated word, in step S55 the speech recognition interface application 12 determines the identified tag for the selected word using the link data 25 and the speech recognition output data. The audio component is then retrieved from the speech recognition run time created files in the temporary directory view the speech recognition engine application 11 in step S56 and in step S57 the audio component is played back via the speech recognition engine application 11. In step S55, once the identified tag has been determined, in addition to retrieval of the audio component, the alternative words from the speech recognition output data in step S58 is obtained and the choice list is built and displayed on the display in step S59. In step S60 a user can select an alternative word from the choice list, input a new word, default back to the original word or cancel if the original word is correct or the word was selected for correction in error. If a user cancels the

operation in step S60a the process proceeds to determine whether more corrections are required. If the user does not cancel the operation in step S61 the displayed document is updated and in step S62 the corrected word and the corresponding identifier flag is sent to the speech recognition engine application 11. In step S63 the speech recognition engine application 11 updates the user's acoustic model within the user model 21. In step S64 the link data is updated, e.g. if the correct word has more characters in it than the replaced word, the character position of all subsequent words will change and thus the link data will need to be updated. In step S65, if more corrections are required the user will in step S50 select another word for correction and repeat the process. Otherwise the correction process is finished and terminates in step S66.

Referring now to figure 8b, this diagram illustrates a method of automatically detecting possible recognition errors in the text. In step S70 the user selects a threshold score to be used to detect possible recognition errors. In step S71 the document or selected text is scanned to compare the threshold score with the score for each of the words. In step S72 the document is scanned to compare the threshold score with the score for the next word. If in step S72 it is found that the score for the word is greater than the threshold, the process proceeds to step S85 where it is determined whether the

end of the document has been reached. If it is not the end of the document then the process returns to step S71 to compare the score for the next word with the threshold score. If in step S72 it is determined that the score for the word is less than the threshold score, the word is highlighted on the display in step S73. In step S74 the speech recognition interface application 12 determines the word location in the text and in step S75 the identifier tag for the word is determined. In step S76 the audio component is retrieved from the speech recognition run time created files in the temporary directory via the speech recognition engine application 11 for playback of the audio component via the speech recognition engine application 11 in step S77. Once the identifier tag is determined in step S75, in step S78 the alternative words for the word having the score less than the threshold is obtained from the output of the speech recogniser engine application 11. In step S79 a choice list is built and displayed on the display. The choice list comprises the list of alternative words displayed alphabetically. In step S80 a user can select an alternative word from the choice list, input a new word, default back to the original word, or cancel if the original word is thought to be correct. If a user cancels the operation in step S80a, the process proceeds to step S85 to determine whether the end of the document or selected text has been reached. If the user does not

cancel the operation, in step S81 the displayed document is updated and in step S82 the corrected word and identifier flag is sent to the speech recogniser engine application 11. In step S83 the speech recognition engine application 11 updates the user's acoustic model in the user model 21. In step S84 the link data is updated, e.g. if the correct word contains more or less than characters than the original word, the character number indicating the position of the first character of all of the following words will change and thus the link data for these words must be updated. In step S85 it is determined whether the end of the document, or the selected text, has been reached. If so, the process is terminated in step S86, otherwise the process returns to step S71 to continue scanning the document or selected text.

Thus in the process described with reference to figures 5 to 8, the user is able to harness the output of the speech recognition engine to maintain links between the words in the text and the corresponding audio components in the audio data even if the words are moved or are dispersed with non dictated text or text which has been dictated at some other time. Link data effectively acts as a pointer between the position of the text in the document and the position of the corresponding audio component in the audio data. In this way the dictated text can be ordered in any way and mixed with non

dictated text without losing the ability to play back the audio components when selected by a user.

Also, since not only audio data but also the link data is stored in non-volatile storage such as the disk storage 15, the user is able to reopen a document and play back the corresponding audio data. This enables a user to dictate a document and store it without correction thereby allowing correction at a later date, i.e. delaying the correction. When the document link data and audio data is read, the system returns to a state as if the text had just been dictated. The text can be corrected and the corrections can be fed back to the speech recognition engine to update the user model 21.

Referring now to figure 9, there is illustrated a flow diagram illustrating the feature of another aspect of the present invention. In figure 9, many steps are the same as those illustrated in figure 5 and thus the same references are used. In this aspect of the present invention, when audio data is associated with a document (S6) and a user selects to read audio data (step S8), the system determines whether there are any audio messages associated with the document in step S90.

If there are no audio messages associated with a document the process proceeds to step S9 where the document and link data is read and the speech recognition run time created files are copied from the user's

directory to the temporary directory and the system proceeds as described with regard to Figure 5. If however there are one or more audio messages associated with the document, the user is given the option to select the audio message which is to be played in step S91. If an audio message is not to be played then the process proceeds to step S9. If however the user selects to play a selected audio message, in step S92 the selected audio message is retrieved from the speech recognition run time created files via the speech recognition engine applications 11 and in step S93 the selected audio message is played via the speech recognition engine application 11. The process then proceeds to step S9 as described with reference to Figure 5. Although Figure 9 illustrates the audio note only being playable at a particular time, an audio note can be played at any time during the creation of a document or after a document has been read.

In Figure 10 there is illustrated a procedure for dictating one or more audio messages which can be carried out at any time. In step S95 the user can elect whether or not to dictate an audio message to be associated with a document to be created. If no audio message is to be created the process terminates in step S99b. If an audio message is to be created in step S96 the dictation of the audio message is initiated and in step S97 the audio message is stored in the speech recognition run time

files. In step S98 it is determined whether the dictation of the audio message has finished and if not the process returns to step S96. If the audio message has finished in step S99 the link data is updated to indicate that the document includes an audio message and in step S99a another audio message can be selected to be dictated and the process returns to step S96. Otherwise the process can be terminated in step S99b.

This aspect of the present invention illustrated in Figures 9 and 10 allows for a user to dictate one or more messages which is stored in association with a document. During the dictation of an audio message no recognised text is input to the text processor application 13. This is achieved in the specific embodiment by failing to pass the text to the text processor application 13. This could alternatively be achieved by disabling the recognition capability of the speech recogniser engine application 11 so that only the audio data is stored.

In the specific example the audio message merely comprises a normal audio data file which has the speech recognition data of Figure 3 in corresponding run time files and which is ignored.

As can be seen with regard to Figure 9, when a user opens a document the link data is examined to determine whether there are any audio messages associated with a document and if so an option is displayed to allow the user to select and play a message. If the user selects

to play the message the link data identifies the audio data file containing the audio message which is retrieved and played back via the speech recognition engine 11.

This aspect of the present invention can be used without the features of correcting the user model and can in its simplest form comprise a method of recording and digitising audio messages and storing the audio messages with a document which could simply be created in a conventional manner without involving speech recognition. The audio message allows for instructions or reminding information to be attached to a document in audio form.

Another aspect of the present invention will now be described with reference to Figures 11 to 15. In this aspect of the present invention the correction of the incorrectly recognised words in a dictated passage of text can be carried out on a machine which is separate to the machine containing the speech recognition engine 11 and user model 21. In Figure 11 there is illustrated a network of author work stations 100a, 100b and 100c which comprise the system as described with regard to Figures 1 to 10. The author work stations 100a, 100b and 100c are connected via a network 101 under the control of a network server 102 to an editor work station 103. The network 101 can comprise any conventional computer network such as an ethernet or token ring.

Although in Figure 11 access to the files of the author work stations is achieved via the network 101, any

method of obtaining copies of the documents, associated link data files, and associated speech recognition run time created files can be used. For instance, the documents could be transferred by copying the relevant files on to a computer readable medium such as a floppy disc which can be read by the editor work station and amended. Also correction files (to be explained hereinafter) can be stored on the disc and the disc can be re-read by the author work station for updating of the user model 21 by the speech recognition engine application 11. Further, although three other work stations and a single editor work station are illustrated any number can be used on the network.

Figure 12 illustrates the architecture of the editor work station 103. Like reference numerals in Figure 12 to the reference numerals of Figure 2 represent like components. In the editor work station 103 there is no user model 21, language model 22, dictionary 23 or SR output data 24 in the memory 20. Also the processor 10 does not include the speech recognition engine application 11 and the speech recognition interface application 12 is replaced with the correcting application 12a. In the disk storage 15 there is no partition of the disk into the temporary directory and the user's directory. The documents can however be stored locally into a disk storage 15. The editor work station differs from the author work station further in

that there is no input/output device 19, digital signal processor 18, and analogue to digital converter 17, audio input device 16, and digital to analogue converter 27. Instead the audio output device 28 (loudspeaker or loudspeakers) receives its output from a conventional multimedia sound card 19a.

The editor work station 103 is also provided with a network card 200 to interface the editor work station 103 with the network 101 to allow for the document, link data and speech recognition run time created files to be read from a correspondence path. Of course, although not illustrated in Figure 2, the author work station 100a, 100b and 100c will include a similar network card 200 in this embodiment.

Figure 13 is a flow diagram of the operation of the editor work station in accordance with the specific embodiment of the present invention. In step S100 the word processor application and a correction application is loaded. The correction application comprises a modified form of the speech recognition interface application. In step S101 the user selects a correspondence path, a user path and a document for correction. The correspondence path is the directory in which the user has saved the document, the link data file, and the speech recognition run time created files. The user path is the directory in which the speech recognition data, specifically the user model 21, is

stored. In step S102 the document and link data file is read. In step S102 the document and link data file can simply be read over the network or it can be copied so that the editor work station 103 has a local copy. If a local copy is made, it is important that when corrections are made the corrected document is stored in the correspondence path together with the amended link data file. In step S103 the link data determines whether there are any audio messages associated with the read document. If there are no audio messages the process proceeds to step S104 for the correction of the document. If an audio message is present in step S105 the user is given an option to select the audio message for playing. If an audio message is not to be played the process proceeds to step S104. If an audio message is to be played the selected audio message is retrieved from the speech recognition run time created files in step S106 and in step S107 the selected audio message is converted to a conventional sound format, e.g. .WAV. In step S108 the audio message is then played through the conventional sound card 19a and loud speakers 28 and the process then proceeds to step S104. Once the document has been corrected, the details of which will be described in more detail hereinafter, in step S109 the editor is given the option as to whether to update the user's contextual model. If the editor does not wish to update the user's contextual model the process proceeds to step S111 where

the editor is given the option as to whether or not to save the document. If the user's contextual model is to be updated in step S110 the user selects text containing corrections whereupon context update parameters are stored in a contextual correction file in the user path. The contextual update parameters include the number of corrected words and a list of the corrected words. The process then proceeds to step S111. If the document is to be saved, in step S112 the document and associated link data is stored in the correspondence path and in step S113 the editor is given the option as to whether to store a copy locally in the editor work station 103 in step S114. In step S115 the editor can then either exit the word processor, in which case the process terminates in step S116, or select another document by returning to step S101.

Referring now to Figure 14a, this document is a flow diagram of the method of manually correcting the document corresponding to step S104 of Figure 13. In step S120 the editor selects a word for correction and in step S121 the word is highlighted on the display. In step S122 the correction application determines the word location in the text and in step S123 it is determined whether the word is a dictated word or not by comparing the word location with the link data 25. If the word is not a dictated word a message is displayed informing the editor that the word is not a dictated word in step S124 and in

step S135 the system awaits further corrections. If the word is a dictated word in step S125 the identified tag is determined. In step S126 the audio component from the speech recognition run time created file is retrieved from the correspondence path and the audio component corresponding to the selected word is converted to a conventional audio format (.WAV) in step S127. The audio component is then played back using the conventional multimedia sound card and loudspeakers in step S128.

Once the identified tag is determined in step S125 the alternative words are read from the speech recognition run time created files in the correspondence path in step S129 and in step S130 a choice list is built and displayed. The choice list comprises the alternative words listed alphabetically for ease of use. In step S131 the editor can select an alternative word from the choice list, input a new word, default back to the original word, or cancel if the original word is considered to be correct or the editor incorrectly selected the word. If an editor cancels the operation in step S131a the process proceeds to step S135 to determine whether more corrections are required. If the user does not cancel the operation, in step S132 the displayed document is updated and in step S133 the corrected word and identifier flag is stored in a word correction file in the user path. In step S134 the link data is updated e.g. if the correct word is of different

length to the replaced word, the character number identifying the position of the first character of each of the proceeding words will be changed and thus the link data for all of the following words must be changed. In step S135, if the user makes no more corrections, the process ends at step S136 otherwise the user can select another word in step S120.

Figure 14b is a flow diagram of an automatic method of correcting recognition errors corresponding to the correction step S104 in Figure 13. In step S140 the editor can select the desired threshold score for the automatic correction process. In step S141 the document or selected text is scanned to compare the score of the next word with the threshold score. In step S142 if the score for the word is greater than the threshold, in step S155 it is determined whether it is the end of the document or selected text and if it is the process terminates in step S156. Otherwise the scanning of the document in step S141 continues for each word in the selected text or until the end of the document is reached. If in step S142 it is determined that the score for a word is less than the threshold an in step S143 the word is highlighted on the display and in step S144 the word location in the text is determined. In step S145 the identifer tag for the word is determined from the link data 25 and in step S146 the audio component is retrieved from the SR run time created files. In step

S147 the audio component is converted to a standard audio format (.WAV format) and in step S148 the audio component is played back using the conventional multimedia sound card 19a and loudspeakers 28.

When the identifer tag is determined for the word in step S145 in step S149 the alternative words from the speech recognition run time created files can be read in the correspondence path and in step S150 a choice list can be built and displayed. The choice list comprises a list of the alternative words in alphabetical order. In step S151 the editor can select an alternative word from the choice list, input a new word, default back to the original word, or cancel if it is considered that the original word was correct. If the editor cancels the operation in step S151 the process proceeds to step S155 to determine whether the end of the document or selected text has been reached. If the editor does not cancel the operation, in step S152 the displayed document is updated and in step S153 the corrected word and identifer flag are stored in a word correction file in the user path. In step S154 the link data 25 is updated e.g. if the correct word has a different length to the original word the position of the following words will change and thus the link data needs to be updated. In step S155 it is determined whether it is the end of the document, or selected text, and if so the process terminates in step S156.

Referring now to Figure 15, this is a flow diagram of the additional steps which are carried out at a networked author work station when the speech recognition engine application and the speech recognition interface application is loaded. In step S160 the speech recognition interface application detects whether there are any word correction files or contextual correction files present in the user path. If no correction files are detected at present then the process terminates in step S161 allowing the user to continue to step S2 in Figures 5 or 9. If correction files are detected to be present in step S160 the author is given the option as to whether to carry out updating of the user model 21 at this time for the selected correction files in step S162. If no updating is to be carried out for the selected correction files the process proceeds to step S167 to determine if there are more correction files present. If the author selects to carry out the updating of the user model 21 using the selected correction files, in step S163 the associated word and/or contextual correction files are read from the user path. In step S164 the speech recognition run time created files are copied from the correspondence path to the temporary directory and in step S165 the word and contextual update parameters are sent to the speech recognition engine application 11 by the speech recognition interface application 12. In step S166 the read correction files

are then deleted in the user path. In step S167 it is then determined whether there are any more correction files present in the user path and if so the user is given the option as to whether to update using these files in step S162. If in step S167 there are no more correction files present then the process terminates in step S161 allowing the user to proceed to step S2 in Figures 5 or 9.

Although in step S162 the author can select each associated word and contextual correction file for updating, the author may also be given the opportunity to elect for the updating to be carried out for all of the correction files present in the user path.

This aspect of the present invention illustrated in Figures 11 to 15 allows an author to dictate documents, save them and delegate correction to an editor by a separate machine. The corrections made by the editor are then fed back to update the author's user model to increase the accuracy of the speech recognition thereafter. However, since the author's user model is not copied, there is no danger of there being more than one copy of the user model whereby one of the copies could be out of date. Also, since the editor does not have access to the author's user model, the corrections being carried out by the editor does not prevent the author from continuing to use the speech recognition engine application which requires access to the user

model. By delegating the correction to the editor whereby updates are generated in files, dictation by the author and correction by the editor can be carried out in parallel.

The delegated correction feature is enhanced by the provision of the audio note capability allowing an author to dictate instructions to the editor to be attached to the document to be edited. The audio message capability can not only be used in conjunction with the delegated correction facility, but can also be used on its own simply to provide audio messages with a document.

The delegated correction system also provides a cost reduction for users since the editor need not be supplied with the speech recognition software and system components. The editor work station 103 can simply comprise a standard multimedia PC. It is of course possible to provide a plurality of such editor work stations in the network to serve any number of author work stations.

The delegated correction system can also operate without a network by physically moving files between the author and editor work stations on computer readable storage media such as floppy disks.

Although in the embodiments described hereinabove word processing is described as occuring after dictation, word processing of the document can take place at any time.

Further, although in the embodiments the recording and playing of audio message is described as occuring at specific points in the process they can be recorded or played at any time.

What has been described hereinabove are specific embodiments and it would be clear to a skilled person in the art that modifications are possible and the present invention is not limited to the specific embodiments.

CLAIMS

1. Data processing apparatus comprising

input means for receiving recognition data and corresponding audio data from a speech recognition engine, said recognition data including a string of recognised characters and audio identifiers identifying audio components corresponding to a character component of the recognised characters;

storage means for storing said audio data received from said input means;

processing means for receiving and processing the input recognised characters to replace, insert and/or move characters in the recognised characters and/or to position the recognised characters;

link means for forming link data linking the audio identifiers to the character component positions in the character string even after processing;

display means for displaying the characters being processed by the said processing means;

user operable selection means for selecting characters in the displayed characters for audio playback, where said link data identifies any selected audio components, if present, which are linked to the selected characters; and

audio playback means for playing back the selected audio components in the order of the character component

positions in the character string.

2. Data processing apparatus as claimed in claim 1 wherein said storage means stores the characters, the link data and the audio data, and storage reading means for reading the stored characters into said processing means and for reading the stored link data for use by said processing means and said link means, whereby said user operable selection means can select displayed characters for audio playback and said audio playback means reads and plays back the audio components corresponding to the selected characters.

3. Data processing apparatus as claimed in claim 1 or claim 2 including user operable correction means for selecting and correcting any displayed recognised characters which have been incorrectly recognised, correction audio playback means for controlling said audio playback means to play back any audio component corresponding to the selected characters to aid correction; and speech recognition update means for sending the corrected characters and the audio identifier for the audio component corresponding to the corrected character to the speech recognition engine.

4. Data processing apparatus as claimed in claim 3 wherein said recognition data includes alternative

characters, said display means including means to display a choice list comprising the alternative characters, said selecting and correcting means including means to select one of the alternative characters or to enter a new character.

5. Data processing apparatus as claimed in any preceding claim wherein said link means comprises memory means storing a list of character locations in the character string and positions of the corresponding audio components in the audio data.

6. Data processing apparatus as claimed in claim 5 wherein said character string is formed of a plurality of separately dictated passages of characters, the apparatus including audio storage means storing said audio data for each dictated passage of characters in a separate file, and said memory means storing a list identifying the files and positions in the files of the audio components in said audio data corresponding to the word locations in the character string.

7. Data processing apparatus as claimed in any preceding claim wherein said recognition data includes recognition status indicators to indicate whether each recognised character is a character finally selected as recognised by said speech recognition engine or a

character which is the most likely at that time but which is still being recognised by said speech recognition engine, the apparatus including status detection means for detecting said recognition status indicators, and display control means to control said display means to display characters which are still being recognised differently to characters which have been recognised, said link means being responsive to said recognition status indicators to link the recognised characters to the corresponding audio component in the audio data.

8. Data processing apparatus as claimed in any preceding claim including contextual update means operable by a user to select recognised characters which are to be used to provide contextual correcting parameters to said speech recognition engine, and to send said contextual correcting parameters to said speech recognition engine.

9. Data processing apparatus as claimed in any preceding claim wherein said recognition data includes a likelihood indicator for each character in the character string indicating the likelihood that the character is correct, and said link means stores the confidence indicators, the apparatus including

automatic error detection means for detecting possible errors in recognition of characters in the

recognised characters by scanning the likelihood indicators in said link means for the recognised characters and detecting if the likelihood indicator for a character is below a threshold, whereby said display means highlights the character having a likelihood indicator below the likelihood threshold;

user operable selection means for selecting a character to replace an incorrectly recognised character highlighted in the recognised characters; and

correction means for replacing the incorrectly recognised character with the selected character to correct the recognised characters.

10. Data processing apparatus as claimed in any preceding claim including

file storage means for storing the recognised characters in a file;

means for selectively disabling one of the receipt of the recognised characters by said processing means and the recognition of speech by said speech recognition engine for a period of time, means for storing the audio data for the period of time in said storage means as an audio message associated with the file; and

storage reading means for reading said file for input to said processing means, and for reading said audio message for playback by said audio playback means.

11. Data processing apparatus as claimed in claim 10 wherein said storage reading means is controllable by a user to read said audio message at any time after said file has been input to said processing means until said processing means is no longer processing said file.

12. Data processing apparatus as claimed in any preceding claim wherein said user operable selection means is operative to allow a user to select to playback the audio data for the most recent passage of dictated characters, or to select characters and play back the corresponding audio components.

13. A data processing network comprising

data processing apparatus as claimed in claim 1 including storage means for storing the characters, the link data and the audio data; and

an editor work station connected to said data processing apparatus via a network, said editor work station comprising

data reading means for reading the characters, link data, and audio data from said data processing apparatus over the network;

editor processing means for processing the characters;

editor link means for linking the audio data to the character component position using the link data;

editor display means for displaying the characters being processed;

editor correction means for selecting and correcting any displayed characters which has been incorrectly recognised; editor audio playback means for playing back any audio component corresponding to the selected characters to aid correction;

editor speech recognition update means for storing the corrected characters and the audio identifier for the audio component corresponding to the corrected character in a character correction file; and

data uploading means for uploading the character correction file to said data processing apparatus for later updating of models used by said speech recognition engine;

said data processing apparatus including correction file reading means for reading said character correction file to pass the data contained therein to said speech recognition engine.

14. A data processing method as claimed in claim 13 wherein said recognition data includes alternative characters, said editor display means including means to display a choice list comprising the alternative characters, said editor correcting means including means to select one of the alternative characters or to enter a new character.

15. A data processing network as claimed in claim 13 or claim 14 including editor contextual update means operable by a user to select recognised characters which are to be used to provide contextual correcting parameters to said speech recognition engine of said data processing apparatus, and to store said contextual correcting parameters in a contextual correction file;

said data uploading means being responsive to the contextual correction file to upload the contextual correction file to said data processing apparatus for later updating of models used by said speech recognition engine;

said correction file reading means of said data processing apparatus being responsive to the contextual correction file to read the contextual correction file to pass the data contained therein to said speech recognition engine.

16. A data processing network as claimed in any one of claims 13 to 15 wherein said recognition data includes a likelihood indicator for each character in the character string indicating the likelihood that the character is correct, and said link data includes the indicators, said editor work station including editor automatic error detection means for detecting possible errors in recognition of characters in the recognised characters by scanning the likelihood indicators in said

data for the characters and detecting if the likelihood indicator for a character is below a likelihood threshold, whereby said editor display means highlights characters having a likelihood indicator below the likelihood threshold;

editor selection means for selecting a character to replace an incorrectly recognised character highlighted in the text; and

editor correction means for replacing the incorrectly recognised character with the selected character to correct the recognised characters.

17. A data processing network as claimed in any one of claims 13 to 16 wherein said data processing apparatus includes file storage means for storing the recognised characters in a file; means for selectively disabling one of the receipt of the recognised characters by said processing means and the recognition of speech by said speech recognition engine for a period of time, means for storing the audio data for the period of time in said storage means as an audio message associated with the document; and storage reading means for reading said document for input to said processing means, and for reading said audio message for playback by said audio playback means; said editor work station including audio message reading means for reading over the network the audio message associated with characters being processed

by said editor processing means for playback by said editor audio playback means.

18. A data processing network as claimed in claim 17 wherein said audio message reading means is controllable by a user to read said audio message at any time the associated characters are being processed by said editor processing means.

19. An editor work station for use with the data processing network as claimed in any one of claims 13 to 18, said editor work station comprising:

data reading means for reading the characters, link data, and audio data from said data processing apparatus over the network;

editor processing means for processing characters;

editor link means for linking the audio data to the character component position using the link data;

editor display means for displaying the characters being processed;

editor correction means for selecting and correcting any displayed characters which have been incorrectly recognised; editor audio playback means for playing back any audio component corresponding to the selected characters to aid correction;

editor speech recognition update means for storing the corrected character and the audio identifier for the

audio component corresponding to the corrected character in a character correction file; and

data uploading means for uploading the character correction file to said data processing apparatus for later updating of models used by said speech recognition engine.

20. An editor work station as claimed in claim 19 wherein said recognition data includes alternative characters, said editor display means including means to display a choice list comprising the alternative characters, said editor correcting means including means to select one of the alternative characters or to enter a new character.

21. An editor work station as claimed in claim 19 or claim 20 including editor contextual update means operable by a user to select recognised characters which are to be used to provide contextual correcting parameters to said speech recognition engine of said data processing apparatus, and to store said contextual correcting parameters in a contextual correction file;

said data uploading means being responsive to the contextual correction file to upload the contextual correction file to said data processing apparatus for later updating of models used by said speech recognition engine;

said correction file reading means of said data processing apparatus being responsive to the contextual correction file to read the contextual correction file to pass the data contained therein to said speech recognition engine.

22. An editor work station as claimed in any one of claims 19 to 21 wherein said recognition data includes a likelihood indicator for each character in the character string indicating the likelihood that the character is correct, and said link data includes the indicators, said editor work station including editor automatic error detection means for detecting possible errors in recognition of characters in the recognised characters by scanning the likelihood indicators in said data for the characters and detecting if the likelihood indicator for a character is below a likelihood threshold, whereby said editor display means highlights characters having a likelihood indicator below the likelihood threshold;

editor selection means for selecting a character to replace an incorrectly recognised word highlighted in the character string; and

editor correction means for replacing the incorrectly recognised character with the selected character to correct the recognised text.

23. A data processing method comprising the steps of:

receiving recognition data and corresponding audio data from a speech recognition engine, said recognition data including recognised characters and audio identifiers identifying audio components corresponding to text components in the recognised text;

inputting the recognised characters to a processor for the processing of the characters to at least one of replace, insert and move words in the character, position the character, and format the characters;

forming link data linking the audio identifiers to the character component positions in the characters even after processing;

displaying the characters input to the processor;

selecting displayed characters for audio playback, whereby said link data identifies any selected audio components, if present, which are linked to the selected characters; and

playing back the selected audio components in the order of the character component positions in the characters.

24. A method as claimed in claim 23 wherein the characters, the link data and the audio data is stored, the method including the step of reading the stored characters into the processor and reading the stored link data, whereby any of the read characters can be selected

for audio playback, the read back data links the selected read characters to any corresponding stored audio data, and corresponding audio data is read and played back.

25. A method as claimed in claim 23 or claim 24 including the steps of selecting any displayed characters which has been incorrectly recognised, playing back any audio component corresponding to the selected characters to aid correction, correcting the incorrectly recognised characters, and sending the corrected characters and audio identifier for the audio component to the corrected character to the speech recognition engine.

26. A method as claimed in claim 25 wherein said recognition data includes alternative characters, the method including the step of displaying a choice list when any displayed characters have been selected for correction, said choice list comprising said alternative characters; and

said correcting step comprises selecting one of the alternative characters or inputting a new character.

27. A method as claimed in any one of claims 23 to 26 wherein said link data comprises a list of character locations in the characters and positions of the corresponding audio components in the audio data.

28. A method as claimed in claim 27 wherein said text is formed of a plurality of separately dictated passages of characters, the method including the steps of storing said audio data for each dictated passage of characters in separate files, said link data including a list identifying the files and positions in the files of the audio components in said audio data corresponding to the word locations in the characters.

29. A method as claimed in any one of claims 23 to 28 wherein said recognition data includes recognition status indicators to indicate whether each recognised character is a character finally selected as recognised by said speech recognition engine or a character which is the most likely at that time but which is still being recognised by said speech recognition engine, the method including the steps of detecting said recognition status indicators, displaying characters which are still being recognised differently to the characters which have been recognised, and forming said link data by linking the positions of the recognised characters in the characters to the positions of the corresponding audio components in the audio data.

30. A method as claimed in claim 25 or claim 26 including the steps of selecting recognised characters which are to be used to provide contextual correcting

parameters to said speech recognition engine, and sending the contextual correcting parameters to said speech recognition engine.

31. A method as claimed in any one of claims 23 to 30 wherein said recognition data includes a likelihood indicator for each character in the characters indicating the likelihood that the character is correct, the method including the steps of

detecting possible errors in recognition of characters in the characters by scanning the likelihood indicators for the characters, and detecting if the likelihood indicator for a character is below a likelihood threshold;

highlighting the character having a likelihood indicator below the likelihood threshold;

if the highlighted character is an incorrectly recognised character, selecting a character to replace an incorrectly recognised character highlighted in the characters; and

replacing the incorrectly recognised character with the selected character to correct the characters.

32. A method as claimed in any one of claims 23 to 31 including the steps of storing the characters as a file;

selectively disabling one of the importation of recognised characters into the processor and the

recognition of speech by said speech recognition engine for a period of time;

storing the audio data for the period of time as an audio message associated with the file;

at a later time, reading said file for input to the processor; and

allowing a user to select whether to read and playback said audio message associated with said file.

33. A method as claimed in claim 32 wherein said audio message can be read and played back at any time said file is open in the processor.

34. A method as claimed in any one of claims 23 to 33 including the step of allowing a user to select to playback the audio data for the most recent passage of dictated characters.

35. A method of processing data over a network comprising the steps of:

at an author work station, carrying out the method as claimed in claim 23 wherein the characters, the link data and the audio data is stored; and

at an editor work station linked to said author work station by said network, reading the stored characters, link data and audio data from the author work station over said network;

inputting the characters into a processor;

linking the audio data to the character component positions using the link data;

displaying the characters being processed;

selecting any displayed characters which have been incorrectly recognised;

playing back any audio component corresponding to the selected characters to aid correction;

correcting the incorrectly recognised characters;

storing the corrected characters and the audio identifier for the audio component corresponding to the corrected character in a character correction file; and

uploading the character correction file over the network to the author work station for later updating of models used by said speech recognition engine;

wherein, at a later time, said character correction file is read at said author work station to pass the data contained therein to said speech recognition engine for updating of said models.

36. A method as claimed in claim 35 wherein said recognition data includes alternative characters, the correcting step at said editor work station, comprising the steps of displaying a choice list comprising the alternative characters, and selecting one of the alternative characters or entering a new character.

37. A method as claimed in claim 35 or claim 36 including the steps at said editor work station of selecting recognised characters which are to be used to provide contextual correcting parameters to said speech recognition engine at said author work station;

storing said contextual correcting parameters in a contextual correction file; and

uploading said contextual correction file over the network to said author work station for later updating of models used by said speech recognition engine; and

at said author work station, at a later time, reading the uploaded contextual correction file and passing the data contained therein to said speech recognition engine.

38. A method as claimed in any one of claims 35 to 37 wherein said recognition data includes a likelihood indicator for each character in the characters indicating the likelihood that the character is correct, the method including the steps at said editor work station of

automatically detecting possible errors in recognition of characters by scanning the likelihood indicators for the characters;

detecting if the likelihood indicator for a character is below a likelihood threshold, whereby characters having a likelihood indicator below the likelihood threshold are displayed highlighted;

selecting a character to replace an incorrectly recognised character highlighted in the characters; and

replacing the incorrectly recognised character with the selected character to correct the characters.

39. A method as claimed in any one of claims 35 to 38 wherein the method includes the steps of:

at said author work station, storing the characters as a file;

selectively disabling one of the importation of recognised characters into the processor and the recognition of speech by said speech recognition engine for a period of time;

storing the audio data for the period of time as an audio message associated with the file;

at a later time, reading said file for input to the processor; and,

at said editor work station, reading over the network the audio message associated with the file being processed by the processor, and playing back the read audio message.

40. A method as claimed in claim 39 wherein the audio message can be read and played back at any time said file is open in the processor.

41. A method as claimed in any one of claims 35 to 40

including the step of allowing a user of the editor work station to playback the audio data for the most recent passage of dictated characters.

42. A data processing network as claimed in any one of claims 13 to 22 comprising a plurality of said data processing apparatus connected to the network, and at least one editor work station, wherein each editor work station can access and edit stored characters and audio data on a plurality of said data processing apparatus.

43. Data processing apparatus comprising

means for receiving recognition data from a speech recognition engine and corresponding audio data; the recognition data including recognised characters;

display means for displaying the recognised characters;

storage means for storing the recognised characters as a file;

means for selectively disabling one of the display and storage of the recognised characters and the speech recognition engine for a period of time; and

means for storing the audio data for the period of time in said storage means as an audio message associated with the file.

44. Data processing apparatus as claimed in claim 43

including reading means for reading the file for display on said display means and for reading said audio message associated with the file; and

audio play back means for playing back the read audio message.

45. Data processing apparatus comprising means for reading a file and associated audio message stored using the data processing apparatus of claim 43, display means for displaying the file, and audio playback means for playing back the audio message.

46. Data processing apparatus comprising

means for receiving data from a speech recognition engine and corresponding audio data, the recognition data including recognised characters;

display means for displaying the recognised characters;

storage means for storing the recognised characters as a file and for storing the corresponding audio data.

47. Data processing apparatus as claimed in claim 46 including reading means for reading the file for display on said display means and for reading the corresponding audio data; and

audio playback means for playing back the read audio data.

48. Data processing apparatus comprising means for reading a file and corresponding audio data stored using the data processing apparatus of claim 46, display means for displaying the file, and audio playback means for playing back the read audio data.

49. Data processing apparatus comprising

means for receiving recognition data from a speech recognition engine and corresponding audio data, said recognition data including recognised characters representing the recognised characters and audio identifier identifying the audio component corresponding to a character in the recognised characters;

storage means for storing said audio data and the recognised characters;

display means for displaying the recognised characters received from said speech recognition means or retrieved from said storage means;

user operable selection and correction means for selecting and correcting any displayed recognised characters;

audio playback means for playing back any audio component corresponding to the selected characters to aid correction; and

speech recognition update means for sending the corrected character and the audio identifier for the audio component corresponding to the corrected character

to the speech recognition engine.

50. Data correction apparatus comprising

means for receiving recognition data from a speech recognition engine, said recognition data including recognised characters representing the most likely characters, and a likelihood indicator for each character indicating the likelihood that the character is correct;

display means for displaying the recognised characters;

automatic error detection means for detecting possible errors in recognition of characters in the recognised characters by scanning the likelihood indicators for the recognised characters and detecting if the likelihood indicator for a character is below a likelihood threshold, whereby said display means highlights at least the first, if any, character having a likelihood indicator below the likelihood threshold;

user operable selection means for selecting a character to replace an incorrectly recognised character highlighted in the recognised characters; and

correction means for replacing the incorrectly recognised character with the selected character to correct the recognised characters.

51. Data processing apparatus as claimed in claim 50 including likelihood threshold adjustment means operable

by a user to adjust and set the likelihood threshold to a desired level.

52. A computer usable medium having computer readable instructions stored therein for causing a processor in a data processing apparatus to process signals defining a string of characters and corresponding audio data to display the characters and selectively play the audio data, the instructions comprising instructions for:

a) causing the processor to receive the signals from a speech recognition engine, the recognition signals including recognised characters and audio identifier identifying the audio components corresponding to character components in the recognised characters;

b) causing the processor to process the signals to manipulate the characters;

c) causing the processor to process the signals to form link data linking the audio identifier to the character component positions in the character string;

d) causing the processor to generate an image of the characters on a display;

e) causing the processor to receive a selection signal generated by a user and to identify any audio components corresponding to the selected characters; and

f) causing the processor to send the identified audio components in the order of the character component positions in the characters to an audio play back device.

53. A computer usable medium having computer readable instructions stored therein for causing the processor in a data processing apparatus to process signals defining a string of characters and audio data to store the characters and the audio data, the instructions comprising instructions for

a) causing the processor to receive the signals from a speech recognition engine;

b) causing the processor to generate an image of the characters on a display;

c) causing the processor to store the characters as a file;

d) causing the processor to selectively disable one of the display and storage of the characters and the speech recognition engine for a period of time; and

e) causing the processor to store the audio signal for the period of time as an audio message associated with the file.

54. A computer usable medium as claimed in claim 53 including instructions for

a) causing the processor to read the stored characters and audio signal;

b) causing the processor to generate an image of the characters for display; and

c) causing the processor to send the audio signal to an audio play back device.

55. A computer usable medium having computer readable instructions stored therein for causing a processor in a data processing apparatus to process signals defining a string of characters and corresponding audio data to store the characters and the audio data, the instructions comprising instructions for:

a) causing the processor to receive the signals from a speech recognition engine;

b) causing the processor to generate an image of the characters for display; and

c) causing the processor to store the characters as a file and to store the corresponding audio signal.

56. A computer usable medium having computer readable instructions stored therein for causing a processor in a data processing apparatus to process signals defining a string of characters and corresponding audio data from a speech recognition engine to update the models used by speech recognition engine, the instructions comprising instructions for:

a) causing the processor to receive the characters, audio data, and audio identifiers from the speech recognition engine, said audio identifier identifying audio components corresponding to components in the characters;

b) causing the processor to store the audio data and the characters, in a storage device;

c) causing the processor to generate an image for display of the characters received from the speech recognition engine or retrieved from the storage device;

d) causing the processor to receive a selection signal generated by a user to select characters which have been incorrectly recognised by the speech recognition engine;

e) causing the processor to retrieve any audio component from the storage device corresponding to the selected characters and to send the retrieved audio to an audio play back device;

f) causing the processor to receive corrected characters input by a user and to replace the incorrect characters with the corrected characters; and

g) causing the processor to send the corrected characters and the audio identifier for the audio component corresponding to the corrected characters to the speech recognition engine for the correction of models used by the speech recognition engine.

57. A data processing apparatus as claimed in claim 1 including storage means for storing the characters, the link data and the audio data.

58. An editor work station for editing the text stored by the data processing apparatus of claim 57, the editor work station comprising

reading means for reading the characters, link data, and audio data;

editor processing means for processing the characters;

editor link means for linking the audio data to the character component positions using the link data;

editor display means for displaying the characters being processed;

editor correction means for selecting and correcting any displayed characters which have been incorrectly recognised;

editor audio playback means for playing back any audio component corresponding to the selected characters to aid correction;

editor speech recognition update means for storing the corrected characters and the audio identifier for the audio component corresponding to the corrected characters in a character correction file for later reading by the speech recognition engine of said data processing apparatus to update models used by said speech recognition engine; and

writing means for storing the correct characters and link data and the audio data.

Amendments to the claims have been filed as follows

1. Date processing apparatus comprising

input means for receiving recognition data and corresponding audio data from a speech recognition engine, said recognition data including a string of recognised characters and audio identifiers identifying audio components corresponding to a character component of the recognised characters;

storage means for storing said audio data received from said input means;

processing means for receiving and processing the input recognised characters to at least one of replace, insert, move, and position the recognised characters to form a processed character string;

link means for forming link data linking the audio identifiers to the character component positions in the character string and for updating said link data after processing to maintain character string the link between the audio identifiers and the character component positions in the processed character string;

display means for displaying the characters received by the said processing means;

user operable selection means for selecting characters in the displayed characters for audio playback, where said link data identifies any selected audio components, if present, which are linked to the

selected characters; and

audio playback means for playing back the selected audio components in the order of the character component positions in the character string or the processed character string.

2. Data processing apparatus as claimed in claim 1 wherein said storage means stores the characters, the link data and the audio data, and storage reading means for reading the stored characters into said processing means and for reading the stored link data for use by said processing means and said link means, whereby said user operable selection means can select displayed characters for audio playback and said audio playback means reads and plays back the audio components corresponding to the selected characters.

3. Data processing apparatus as claimed in claim 1 or claim 2 including user operable correction means for selecting and correcting any displayed recognised characters which have been incorrectly recognised, correction audio playback means for controlling said audio playback means to play back any audio component corresponding to the selected characters to aid correction; and speech recognition update means for sending the corrected characters and the audio identifier for the audio component corresponding to the corrected

character to the speech recognition engine.

4. Data processing apparatus as claimed in claim 3 wherein said recognition data includes alternative characters, said display means including means to display a choice list comprising the alternative characters, said selecting and correcting means including means to select one of the alternative characters or to enter a new character.

5. Data processing apparatus as claimed in any preceding claim wherein said link means comprises memory means storing a list of character locations in the character string and positions of the corresponding audio components in the audio data.

6. Data processing apparatus as claimed in claim 5 wherein said character string is formed of a plurality of separately dictated passages of characters, the apparatus including audio storage means storing said audio data for each dictated passage of characters in a separate file, and said memory means storing a list identifying the files and positions in the files of the audio components in said audio data corresponding to the word locations in the character string.

7. Data processing apparatus as claimed in any

preceding claim wherein said recognition data includes recognition status indicators to indicate whether each recognised character is a character finally selected as recognised by said speech recognition engine or a character which is the most likely at that time but which is still being recognised by said speech recognition engine, the apparatus including status detection means for detecting said recognition status indicators, and display control means to control said display means to display characters which are still being recognised differently to characters which have been recognised, said link means being responsive to said recognition status indicators to link the recognised characters to the corresponding audio component in the audio data.

8. Data processing apparatus as claimed in any preceding claim including contextual update means operable by a user to select recognised characters which are to be used to provide contextual correcting parameters to said speech recognition engine, and to send said contextual correcting parameters to said speech recognition engine.

9. Data processing apparatus as claimed in any preceding claim wherein said recognition data includes a likelihood indicator for each character in the character string indicating the likelihood that the

character is correct, and said link means stores the confidence indicators, the apparatus including

automatic error detection means for detecting possible errors in recognition of characters in the recognised characters by scanning the likelihood indicators in said link means for the recognised characters and detecting if the likelihood indicator for a character is below a threshold, whereby said display means highlights the character having a likelihood indicator below the likelihood threshold;

user operable selection means for selecting a character to replace an incorrectly recognised character highlighted in the recognised characters; and

correction means for replacing the incorrectly recognised character with the selected character to correct the recognised characters.

10. Data processing apparatus as claimed in any preceding claim including

file storing means for storing the recognised characters in a file;

means for selectively disabling one of the receipt of the recognised characters by said processing means and the recognition of speech by said speech recognition engine for a period of time, means for storing the audio data for the period of time in said storage means as an audio message associated with the file; and

storage reading means for reading said file for input to said processing means, and for reading said audio message for playback by said audio playback means.

11. Data processing apparatus as claimed in claim 10 wherein said storage reading means is controllable by a user to read said audio message at any time after said file has been input to said processing means until said processing means is no longer processing said file.

12. Data processing apparatus as claimed in any preceding claim wherein said user operable selection means is operative to allow a user to select to playback the audio data for the most recent passage of dictated characters, or to select characters and play back the corresponding audio components.

13. A data processing arrangement comprising

data processing apparatus as claimed in claim 1 including storage means for storing the characters, the link data and the audio data; and

an editor work station comprising

data reading means for reading the characters, link data, and audio data from said data processing apparatus;

editor processing means for processing the characters;

editor link means for linking the audio data to the

character component position using the link data;

editor display means for displaying the characters being processed;

editor correction means for selecting and correcting any displayed characters which have been incorrectly recognised;

editor audio playback means for playing back any audio component corresponding to the selected characters to aid correction;

editor speech recognition update means for storing the corrected characters and the audio identifer for the audio component corresponding to the corrected character in a character correction file; and

data transferring means for transferring the character correction file to said data processing apparatus for later updating of models used by said speech recognition engine;

said data processing apparatus including correction file reading means for reading said character correction file to pass the data contained therein to said speech recognition engine for the updating of the models used by said speech recognition engine.

14. A data processing arrangement as claimed in claim 13 wherein said recognition data includes alternative characters, said editor display means including means to display a choice list comprising the alternative

characters, said editor correcting means including means to select one of the alternative characters or to enter a new character.

15. A data processing arrangement as claimed in claim 13 or claim 14 including editor contextual update means operable by a user to select recognised characters which are to be used to provide contextual correcting parameters to said speech recognition engine of said data processing apparatus, and to store said contextual correcting parameters in a contextual correction file;

said data transfer means being responsive to the contextual correction file to transfer the contextual correction file to said data processing apparatus for later updating of models used by said speech recognition engine;

said correction file reading means of said data processing apparatus being responsive to the contextual correction file to read the contextual correction file to pass the data contained therein to said speech recognition engine.

16. A data processing arrangement as claimed in any one of claims 13 to 15 wherein said recognition data includes a likelihood indicator for each character in the character string indicating the likelihood that the character is correct, and said link data includes the

indicators, said editor work station including editor automatic error detection means for detecting possible errors in recognition of characters in the recognised characters by scanning the likelihood indicators in said data for the characters and detecting if the likelihood indicator for a character is below a likelihood threshold, whereby said editor display means highlights characters having a likelihood indicator below the likelihood threshold;

editor selection means for selecting a character to replace an incorrectly recognised character highlighted in the text; and

editor correction means for replacing the incorrectly recognised character with the selected character to correct the recognised characters.

17. A data processing arrangement as claimed in any one of claims 13 to 16 wherein said data processing apparatus includes file storage means for storing the recognised characters in a file; means for selectively disabling one of the receipt of the recognised characters by said processing means and the recognition of speech by said speech recognition engine for a period of time;

means for storing the audio data for the period of time in said storage means as an audio message associated with the document; and storage reading means for reading said document for input to said processing means, and for

reading said audio message for playback by said audio playback means; said editor work station including audio message reading means for reading the audio message associated with characters being processed by said editor processing means for playback by said editor audio playback means.

18. A data processing arrangement as claimed in claim 17 wherein said audio message reading means is controllable by a user to read said audio message at any time the associated characters are being processed by said editor processing means.

19. An editor work station for use with the data processing arrangement as claimed in any one of claims 13 to 18, said editor work station comprising:

data reading means for reading the characters, link data, and audio data from said data processing apparatus;

editor processing means for processing characters;

editor link means for linking the audio data to the character component position using the link data;

editor display means for displaying the read characters;

editor correction means for selecting and correcting any displayed characters which have ben incorrectly recognised;

editor audio playback means for playing back any

audio component corresponding to the selected characters to aid correction;

editor speech recognition update means for storing the corrected character and the audio identifer for the audio component corresponding to the corrected character in a character correction file; and

data transfer means for transferring the character correction file to said data processing apparatus for later updating of models used by said speech recognition engine.

20. An editor work station as claimed in claim 19 wherein said recognition data includes alternative characters, said editor display means including means to display a choice list comprising the alternative characters, said editor correcting means including means to select one of the alternative characters or to enter a new character.

21. An editor work station as claimed in claim 19 or claim 20 including editor contextual update means operable by a user to select recognised characters which are to be used to provide contextual correcting parameters to said speech recognition engine of said data processing apparatus, and to store said contextual correcting parameters in a contextual correction file;

said data transfer means being responsive to the

contextual correction file to transfer the contextual correction file to said data processing apparatus for later updating of models used by said speech recognition engine;

said correction file reading means of said data processing apparatus being responsive to the contextual correction file to read the contextual correction file to pass the data contained therein to said speech recognition engine.

22. An editor work station as claimed in any one of claims 19 to 21 wherein said recognition data includes a likelihood indicator for each character in the character string indicating the likelihood that the character is correct, and said link data includes the indicators, said editor work station including editor automatic error detection means for detecting possible errors in recognition of characters in the recognised characters by scanning the likelihood indicators in said data for the characters and detecting if the likelihood indicator for a character is below a likelihood threshold, whereby said editor display means highlights characters having a likelihood indicator below the likelihood threshold;

editor selection means for selecting a character to replace an incorrectly recognised word highlighted in the character string; and

editor correction means for replacing the incorrectly recognised character with the selected character to correct the recognised text.

23. A data processing method comprising the steps of:

receiving recognition data and corresponding audio data from a speech recognition engine, said recognition data including recognised characters and audio identifiers identifying audio components corresponding to text components in the recognised text;

storing the audio data;

inputting the recognised characters to a processor for the processing of the characters to at least one of replace, insert, move and position the characters to form a processed character string;

forming link data linking the audio identifiers to the character component positions in the characters and updating said link data after processing to maintain the link between the audio identifiers and the character component positions in the processed character string;

displaying the characters input to the processor;

selecting displayed characters for audio playback, whereby said link data identifies any selected audio components, if present, which are linked to the selected characters; and

playing back the selected audio components in the order of the character component positions in the

character string or the processed character string.

24. A method as claimed in claim 23 wherein the characters, the link data and the audio data is stored, the method including the step of reading the stored characters into the processor and reading the stored link data, whereby any of the read characters can be selected for audio playback, the read back data links the selected read characters to any corresponding stored audio data, and corresponding audio data is read and played back.

25. A method as claimed in claim 23 or claim 24 including the steps of selecting any displayed characters which has been incorrectly recognised, playing back any audio component corresponding to the selected characters to aid correction, correcting the incorrectly recognised characters, and sending the corrected characters and audio identifier for the audio component to the corrected character to the speech recognition engine.

26. A method as claimed in claim 25 wherein said recognition data includes alternative characters, the method including the step of displaying a choice list when any displayed characters have been selected for correction, said choice list comprising said alternative characters; and

said correcting step comprises selecting one of the

alternative characters or inputting a new character.

27. A method as claimed in any one of claims 23 to 26 wherein said link data comprises a list of character locations in the characters and positions of the corresponding audio components in the audio data.

28. A method as claimed in claim 27 wherein said text is formed of a plurality of separately dictated passages of characters, the method including the steps of storing said audio data for each dictated passage of characters in separate files, said link data including a list identifying the files and positions in the files of the audio components in said audio data corresponding to the word locations in the characters.

29. A method as claimed in any one of claims 23 to 28 wherein said recognition data includes recognition status indicators to indicate whether each recognised character is a character finally selected as recognised by said speech recognition engine or a character which is the most likely at that time but which is still being recognised by said speech recognition engine, the method including the steps of detecting said recognition status indicators, displaying characters which are still being recognised differently to the characters which have been recognised, and forming said link data by linking the

positions of the recognised characters in the characters to the positions of the corresponding audio components in the audio data.

30. A method as claimed in claim 25 or claim 26 including the steps of selecting recognised characters which are to be used to provide contextual correcting parameters to said speech recognition engine, and sending the contextual correcting parameters to said speech recognition engine.

31. A method as claimed in any one of claims 23 to 30 wherein said recognition data includes a likelihood indicator for each character in the characters indicating the likelihood that the character is correct, the method including the steps of

detecting possible errors in recognition of characters in the characters by scanning the likelihood indicators for the characters, and detecting if the likelihood indicator for a character is below a likelihood threshold;

highlighting the character having a likelihood indicator below the likelihood threshold;

if the highlighted character is an incorrectly recognised character, selecting a character to replace an incorrectly recognised character highlighted in the characters; and

replacing the incorrectly recognised character with the selected character to correct the characters.

32. A method as claimed in any one of claims 23 to 31 including the steps of storing the characters as a file;

selectively disabling one of the importation of recognised characters into the processor and the recognition of speech by said speech recognition engine for a period of time;

storing the audio data for the period of time as an audio message associated with the file;

at a later time, reading said file for input to the processor; and

allowing a user to select whether to read and play back said audio message associated with said file.

33. A method as claimed in claim 32 wherein said audio message can be read and played back at any time said file is open in the processor.

34. A method as claimed in any one of claims 23 to 33 including the step of allowing a user to select to play back the audio data for the most recent passage of dictated characters.

35. A method of processing data comprising the steps of:

at an author work station, carrying out the method

as claimed in claim 23 wherein the characters, the link data and the audio data is stored; and

at an editor work station, obtaining the stored characters, link data and audio data from the author work station;

inputting the characters into a processor;

linking the audio data to the character component positions using the link data;

displaying the characters being processed;

selecting any displayed characters which have been incorrectly recognised;

playing back any audio component corresponding to the selected characters to aid correction;

correcting the incorrectly recognised characters;

storing the corrected characters and the audio identifier for the audio component corresponding to the corrected character in a character correction file; and

transferring the character correction file to the author work station for later updating of models used by said speech recognition engine;

wherein, at a later time, said character correction file is read at said author work station to pass the data contained therein to said speech recognition engine for updating of said models.

36. A method as claimed in claim 35 wherein said recognition data includes alternative characters, the

correcting step at said editor work station, comprising the steps of displaying a choice list comprising the alternative characters, and selecting one of the alternative characters or entering a new character.

37. A method as claimed in claim 35 or claim 36 including the steps at said editor work station of selecting recognised characters which are to be used to provide contextual correcting parameters to said speech recognition engine at said author work station;

storing said contextual correcting parameters in a contextual correction file; and

transferring said contextual correction file to said author work station for later updating of models used by said speech recognition engine; and

at said author work station, at a later time, reading the transferred contextual correction file and passing the data contained therein to said speech recognition engine.

38. A method as claimed in any one of claims 35 to 37 wherein said recognition data includes a likelihood indicator for each character in the characters indicating the likelihood that the character is correct, the method including the steps at said editor work station of

automatically detecting possible errors in recognition of characters by scanning the likelihood

indicators for the characters;

detecting if the likelihood indicator for a character is below a likelihood threshold, whereby characters having a likelihood indicator below the likelihood threshold are displayed highlighted;

selecting a character to replace an incorrectly recognised character highlighted in the characters; and

replacing the incorrectly recognised character with the selected character to correct the characters.

39. A method as claimed in any one of claims 35 to 38 wherein the method includes the steps of:

at said author work station, storing the characters as a file;

selectively disabling one of the importation of recognised characters into the processor and the recognition of speech by said speech recognition engine for a period of time;

storing the audio data for the period of time as an audio message associated with the file;

at a later time, reading said file for input to the processor; and

at said editor work station, reading the audio message associated with the file being processed by the processor, and playing back the read audio message.

40. A method as claimed in claim 39 wherein the audio

message can be read and played back at any time said file is open in the processor.

41. A method as claimed in any one of claims 35 to 40 including the step of allowing a user of the editor work station to play back the audio data for the most recent passage of dictated characters.

42. A data processing arrangement as claimed in any one of claims 13 to 18 comprising a plurality of said data processing apparatus connected to a network, and at least one editor work station, wherein each editor work station can access and edit stored characters and audio data on a plurality of said data processing apparatus.

The Patent Office

92

Application No:	GB 9619932.8	**Examiner:**	John Donaldson
Claims searched:	43 to 45, 53, 54	**Date of search:**	28 October 1996

Patents Act 1977
Further Search Report under Section 17

Databases searched:

UK Patent Office collections, including GB, EP, WO & US patent specifications, in:

UK Cl (Ed.O): G4R(REX, RHA, RHB); G4H(HTAD, HTAT)

Int Cl (Ed.6): G06F 3/00, 3/16, 17/00, 17/20, 17/21, 17/22, 17/24; G10L 3/00, 5/00, 5/06, 7/00, 7/08, 9/00, 9/08, 9/18

Other: Online:WPI

Documents considered to be relevant:

Category	Identity of document and relevant passage	Relevant to claims
A	GB 2088106 A (MARCONI), see page 4, lines 2 to 77	-

X	Document indicating lack of novelty or inventive step	A	Document indicating technological background and/or state of the art.
Y	Document indicating lack of inventive step if combined with one or more other documents of same category.	P	Document published on or after the declared priority date but before the filing date of this invention.
&	Member of the same patent family	E	Patent document published on or after, but with priority date earlier than, the filing date of this application.

The Patent Office

93

Application No: GB 9619932.8
Claims searched: 50, 51

Examiner: John Donaldson
Date of search: 28 October 1996

Patents Act 1977
Further Search Report under Section 17

Databases searched:

UK Patent Office collections, including GB, EP, WO & US patent specifications, in:

UK Cl (Ed.O): G4R(RHA, RHB)

Int Cl (Ed.6): G10L 3/00, 5/00, 5/06, 7/00, 7/08, 9/00, 9/08, 9/18

Other: Online:WPI

Documents considered to be relevant:

Category	Identity of document and relevant passage	Relevant to claims
A	GB 2230370 A (SMITHS), see page 12, lines 14 to 21	-

X	Document indicating lack of novelty or inventive step	A	Document indicating technological background and/or state of the art.
Y	Document indicating lack of inventive step if combined with one or more other documents of same category.	P	Document published on or after the declared priority date but before the filing date of this invention.
&	Member of the same patent family	E	Patent document published on or after, but with priority date earlier than, the filing date of this application.

The Patent Office

9-4

Application No:	GB 9619932.8	**Examiner:**	John Donaldson
Claims searched:	1 to 42, 46 to 49, 52, 55 to 58	**Date of search:**	17 October 1996

Patents Act 1977
Search Report under Section 17

Databases searched:

UK Patent Office collections, including GB, EP, WO & US patent specifications, in:

UK Cl (Ed.O): G4R(REX, RHA, RHB); G4H(HTAD, HTAT)

Int Cl (Ed.6): G06F 3/00, 3/16, 17/00, 17/20, 17/21, 17/22, 17/24; G10L 3/00, 5/00, 5/06, 7/00, 7/08, 9/00, 9/08, 9/18

Other: Online:WPI

Documents considered to be relevant:

Category	Identity of document and relevant passage		Relevant to claims
X	EP 0649144 A1	(I B M), see column 1, lines 10 to 39 column 4, lines 7 to 57, column 6, line 18 to column 7, line 31	1, 2, 5, 12, 23, 24, 27, 34, 46 to 48, 52, 55, 57
X	EP 0077194 A1	(SHARP), see page 6, line 13 to page 7, line 24	46 to 48, 55
X	WO 93/07562 A1	(RIVERRUN), see page 1, lines 15 to 19, page 16, line 17 to page 17, line 23, page 19, lines 12 to 38, page 22, lines 6 to 37, page 24, line 7 to page 25, line 17	1, 2, 5, 12, 23, 24, 27, 34, 46 to 48, 52, 55, 57
X	US 5031113	(HÖLLERBAUER), see column 3, line 1 to column 6, line 24	1 to 3, 5, 12, 23 to 25, 27, 34, 46 to 49, 52, 55 to 58

X	Document indicating lack of novelty or inventive step	A	Document indicating technological background and/or state of the art.
Y	Document indicating lack of inventive step if combined with one or more other documents of same category.	P	Document published on or after the declared priority date but before the filing date of this invention.
&	Member of the same patent family	E	Patent document published on or after, but with priority date earlier than, the filing date of this application.